INTERNATIONAL TABLES FOR CRYSTALLOGRAPHY

Volume B

RECIPROCAL SPACE

Edited by

U. SHMUELI

Contributing authors

E. Arnold: CABM & Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854-5638, USA. [2.3]

M. I. Aroyo: Departamento de Fisíca de la Materia Condensada, Facultad de Cienca y Technología, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain. [1.5]

A. Authier: Institut de Minéralogie et de la Physique des Milieux Condensés, Bâtiment 7, 140 rue de Lourmel, 75015 Paris, France. [5.1]

H. Boysen: Department für Geo- und Umweltwissenschaften, Sektion Kristallographie, Ludwig-Maximilians Universität, Theresienstrasse 41, 80333 München, Germany. [4.2]

G. Bricogne: Global Phasing Ltd, Sheraton House, Suites 14–16, Castle Park, Cambridge CB3 0AX, England, and LURE, Bâtiment 209D, Université Paris-Sud, 91405 Orsay, France. [1.3]

P. Coppens: Department of Chemistry, Natural Sciences & Mathematics Complex, State University of New York at Buffalo, Buffalo, New York 14260-3000, USA. [1.2]

J. M. Cowley:† Arizona State University, Box 871504, Department of Physics and Astronomy, Tempe, AZ 85287-1504, USA. [2.5.1, 2.5.2, 4.3, 5.2]

L. M. D. Cranswick: Neutron Program for Materials Research, National Research Council Canada, Building 459, Chalk River Laboratories, Chalk River, Ontario, Canada K0J 1J0. [3.3.4]

T. A. Darden: Laboratory of Structural Biology, National Institute of Environmental Health Sciences, 111 T. W. Alexander Drive, Research Triangle Park, NC 27709, USA. [3.5]

R. Diamond: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England. [3.3.1, 3.3.2, 3.3.3]

D. L. Dorset: ExxonMobil Research and Engineering Co., 1545 Route 22 East, Clinton Township, Annandale, New Jersey 08801, USA. [2.5.8, 4.5.1, 4.5.3]

F. Frey: Department für Geo- und Umweltwissenschaften, Sektion Kristallographie, Ludwig-Maximilians Universität, Theresienstrasse 41, 80333 München, Germany. [4.2]

C. Giacovazzo: Dipartimento Geomineralogico, Campus Universitario, 70125 Bari, Italy, and Institute of Crystallography, Via G. Amendola, 122/O, 70125 Bari, Italy. [2.2]

J. K. Gjønnes: Institute of Physics, University of Oslo, PO Box 1048, N-0316 Oslo 3, Norway. [4.3]

P. Goodman†: School of Physics, University of Melbourne, Parkville, Australia. [5.2]

R. W. Grosse-Kunstleve: Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mailstop 4-230, Berkeley, CA 94720, USA. [1.4]

J.-P. Guigay: European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble, France. [5.3]

T. Haibach: Laboratory of Crystallography, Department of Materials, ETH Hönggerberg, HCI G 511, Wolfgang-Pauli-Strasse 10, CH-8093 Zurich, Switzerland. [4.6]

S. R. Hall: Crystallography Centre, University of Western Australia, Nedlands 6907, WA, Australia. [1.4]

H. Jagodzinski: Department für Geo- und Umweltwissenschaften, Sektion Kristallographie, Ludwig-Maximilians Universität, Theresienstrasse 41, 80333 München, Germany. [4.2]

R. E. Marsh: The Beckman Institute–139–74, California Institute of Technology, 1201 East California Blvd, Pasadena, California 91125, USA. [3.2]

R. P. Millane: Department of Electrical and Computer Engineering, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. [4.5.1, 4.5.2]

A. F. Moodie: Department of Applied Physics, Royal Melbourne Institute of Technology, 124 La Trobe Street, Melbourne, Victoria 3000, Australia. [5.2]

P. A. Penczek: The University of Texas – Houston Medical School, Department of Biochemistry and Molecular Biology, 6431 Fannin, MSB 6.218, Houston, TX 77030, USA. [2.5.6, 2.5.7]

P. S. Pershan: Division of Engineering and Applied Science and The Physics Department, Harvard University, Cambridge, MA 02138, USA. [4.4]

S. Ramaseshan†: Raman Research Institute, Bangalore 560 080, India. [2.4]

M. G. Rossmann: Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA. [2.3]

D. E. Sands: Department of Chemistry, University of Kentucky, Chemistry–Physics Building, Lexington, Kentucky 40506-0055, USA. [3.1]

M. Schlenker: Laboratoire Louis Néel du CNRS, BP 166, F-38042 Grenoble Cedex 9, France. [5.3]

V. Schomaker†: Department of Chemistry, University of Washington, Seattle, Washington 98195, USA. [3.2]

U. Shmueli: School of Chemistry, Tel Aviv University, 69 978 Tel Aviv, Israel. [1.1, 1.4, 2.1]

† Deceased.

† Deceased.

J. C. H. SPENCE: Department of Physics, Arizona State University, Tempe, AZ 95287-1504, USA. [2.5.1]

W. STEURER: Laboratory of Crystallography, Department of Materials, ETH Hönggerberg, HCI G 511, Wolfgang-Pauli-Strasse 10, CH-8093 Zurich, Switzerland. [4.6]

M. TANAKA: Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Japan. [2.5.3]

L. TONG: Department of Biological Sciences, Columbia University, New York 10027, USA. [2.3]

B. K. VAINSHTEIN†: Institute of Crystallography, Academy of Sciences of Russia, Leninsky prospekt 59, Moscow B-117333, Russia. [2.5.4, 2.5.5, 2.5.6]

M. VIJAYAN: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. [2.4]

D. E. WILLIAMS†: Department of Chemistry, University of Louisville, Louisville, Kentucky 40292, USA. [3.4]

B. T. M. WILLIS: Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, England. [4.1]

A. J. C. WILSON†: St John's College, Cambridge, England. [2.1]

H. WONDRATSCHEK: Institut für Kristallographie, Universität, D-76128 Karlsruhe, Germany. [1.5]

B. B. ZVYAGIN†: Institute of Ore Mineralogy (IGEM), Academy of Sciences of Russia, Staromonetny 35, 109017 Moscow, Russia. [2.5.4]

† Deceased.

† Deceased.

Contents

PAGE

Preface (U. Shmueli) .. xiii

Preface to the second edition (U. Shmueli) .. xiii

Preface to the third edition (U. Shmueli) .. xiv

PART 1. GENERAL RELATIONSHIPS AND TECHNIQUES

1.1. Reciprocal space in crystallography (U. Shmueli) .. 2

1.1.1. Introduction .. 2

1.1.2. Reciprocal lattice in crystallography .. 2

1.1.3. Fundamental relationships .. 3

1.1.4. Tensor-algebraic formulation .. 5

1.1.5. Transformations .. 7

1.1.6. Some analytical aspects of the reciprocal space .. 8

1.2. The structure factor (P. Coppens) .. 10

1.2.1. Introduction .. 10

1.2.2. General scattering expression for X-rays .. 10

1.2.3. Scattering by a crystal: definition of a structure factor .. 10

1.2.4. The isolated-atom approximation in X-ray diffraction .. 10

1.2.5. Scattering of thermal neutrons .. 11

1.2.6. Effect of bonding on the atomic electron density within the spherical-atom approximation: the kappa formalism .. 11

1.2.7. Beyond the spherical-atom description: the atom-centred spherical harmonic expansion .. 12

1.2.8. Fourier transform of orbital products .. 15

1.2.9. The atomic temperature factor .. 17

1.2.10. The vibrational probability distribution and its Fourier transform in the harmonic approximation .. 17

1.2.11. Rigid-body analysis .. 18

1.2.12. Treatment of anharmonicity .. 20

1.2.13. The generalized structure factor .. 22

1.2.14. Conclusion .. 22

1.3. Fourier transforms in crystallography: theory, algorithms and applications (G. Bricogne) .. 24

1.3.1. General introduction .. 24

1.3.2. The mathematical theory of the Fourier transformation .. 24

1.3.3. Numerical computation of the discrete Fourier transform .. 52

1.3.4. Crystallographic applications of Fourier transforms .. 62

1.4. Symmetry in reciprocal space (U. Shmueli) .. 114

1.4.1. Introduction .. 114

1.4.2. Effects of symmetry on the Fourier image of the crystal .. 114

1.4.3. Structure-factor tables .. 117

1.4.4. Symmetry in reciprocal space: space-group tables .. 119

Appendix A1.4.1. Comments on the preparation and usage of the tables .. 122

Appendix A1.4.2. Space-group symbols for numeric and symbolic computations (U. Shmueli, S. R. Hall and R. W. Grosse-Kunstleve) .. 122

Appendix A1.4.3. Structure-factor tables .. 135

Appendix A1.4.4. Crystallographic space groups in reciprocal space .. 162

1.5. Crystallographic viewpoints in the classification of space-group representations (M. I. Aroyo and H. Wondratschek) .. 175

1.5.1. List of abbreviations and symbols .. 175

1.5.2. Introduction .. 175

1.5.3. Basic concepts .. 176

1.5.4. Conventions in the classification of space-group irreps .. 179
1.5.5. Examples and discussion .. 182
1.5.6. Conclusions .. 191
Appendix A1.5.1. Reciprocal-space groups $(\mathcal{G})^*$ .. 192

PART 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

2.1. Statistical properties of the weighted reciprocal lattice (U. Shmueli and A. J. C. Wilson) .. 195
2.1.1. Introduction .. 195
2.1.2. The average intensity of general reflections .. 195
2.1.3. The average intensity of zones and rows .. 196
2.1.4. Probability density distributions – mathematical preliminaries .. 197
2.1.5. Ideal probability density distributions .. 200
2.1.6. Distributions of sums, averages and ratios .. 202
2.1.7. Non-ideal distributions: the correction-factor approach .. 203
2.1.8. Non-ideal distributions: the Fourier method .. 207

2.2. Direct methods (C. Giacovazzo) .. 215
2.2.1. List of symbols and abbreviations .. 215
2.2.2. Introduction .. 215
2.2.3. Origin specification .. 215
2.2.4. Normalized structure factors .. 216
2.2.5. Phase-determining formulae .. 221
2.2.6. Direct methods in real and reciprocal space: Sayre's equation .. 230
2.2.7. Scheme of procedure for phase determination: the small-molecule case .. 231
2.2.8. Other multisolution methods applied to small molecules .. 232
2.2.9. Some references to direct-methods packages: the small-molecule case .. 234
2.2.10. Direct methods in macromolecular crystallography .. 235

2.3. Patterson and molecular replacement techniques, and the use of noncrystallographic symmetry in phasing (L. Tong, M. G. Rossmann and E. Arnold) .. 244
2.3.1. Introduction .. 244
2.3.2. Interpretation of Patterson maps .. 247
2.3.3. Isomorphous replacement difference Pattersons .. 251
2.3.4. Anomalous dispersion .. 255
2.3.5. Noncrystallographic symmetry .. 258
2.3.6. Rotation functions .. 260
2.3.7. Translation functions .. 269
2.3.8. Molecular replacement .. 272
2.3.9. Conclusions .. 275

2.4. Isomorphous replacement and anomalous scattering (M. Vijayan and S. Ramaseshan) .. 282
2.4.1. Introduction .. 282
2.4.2. Isomorphous replacement method .. 282
2.4.3. Anomalous-scattering method .. 284
2.4.4. Isomorphous replacement and anomalous scattering in protein crystallography .. 287
2.4.5. Anomalous scattering of neutrons and synchrotron radiation. The multiwavelength method .. 293

2.5. Electron diffraction and electron microscopy in structure determination (J. M. Cowley, J. C. H. Spence, M. Tanaka, B. K. Vainshtein, B. B. Zvyagin, P. A. Penczek and D. L. Dorset) .. 297
2.5.1. Foreword (J. M. Cowley and J. C. H. Spence) .. 297
2.5.2. Electron diffraction and electron microscopy (J. M. Cowley) .. 299
2.5.3. Point-group and space-group determination by convergent-beam electron diffraction (M. Tanaka) .. 307

2.5.4. Electron-diffraction structure analysis (EDSA) (B. K. Vainshtein and B. B. Zvyagin) .. 356
2.5.5. Image reconstruction (B. K. Vainshtein) .. 361
2.5.6. Three-dimensional reconstruction (B. K. Vainshtein and P. A. Penczek) .. 366
2.5.7. Single-particle reconstruction (P. A. Penczek) .. 375
2.5.8. Direct phase determination in electron crystallography (D. L. Dorset) .. 388

PART 3. DUAL BASES IN CRYSTALLOGRAPHIC COMPUTING

3.1. Distances, angles, and their standard uncertainties (D. E. Sands) .. 404
3.1.1. Introduction .. 404
3.1.2. Scalar product .. 404
3.1.3. Length of a vector .. 404
3.1.4. Angle between two vectors .. 404
3.1.5. Vector product .. 405
3.1.6. Permutation tensors .. 405
3.1.7. Components of vector product .. 405
3.1.8. Some vector relationships .. 405
3.1.9. Planes .. 406
3.1.10. Variance–covariance matrices .. 406
3.1.11. Mean values .. 408
3.1.12. Computation .. 408

3.2. The least-squares plane (R. E. Marsh and V. Schomaker) .. 410
3.2.1. Introduction .. 410
3.2.2. Least-squares plane based on uncorrelated, isotropic weights .. 410
3.2.3. The proper least-squares plane, with Gaussian weights .. 413
Appendix A3.2.1 .. 416

3.3. Molecular modelling and graphics (R. Diamond and L. M. D. Cranswick) .. 418
3.3.1. Graphics (R. Diamond) .. 418
3.3.2. Molecular modelling, problems and approaches (R. Diamond) .. 434
3.3.3. Implementations (R. Diamond) .. 438
3.3.4. Graphics software for the display of small and medium-sized molecules (L. M. D. Cranswick) .. 442

3.4. Accelerated convergence treatment of R^{-n} lattice sums (D. E. Williams) .. 449
3.4.1. Introduction .. 449
3.4.2. Definition and behaviour of the direct-space sum .. 449
3.4.3. Preliminary description of the method .. 449
3.4.4. Preliminary derivation to obtain a formula which accelerates the convergence of an R^{-n} sum over lattice points X(d) .. 450
3.4.5. Extension of the method to a composite lattice .. 452
3.4.6. The case of $n = 1$ (Coulombic lattice energy) .. 453
3.4.7. The cases of $n = 2$ and $n = 3$ .. 454
3.4.8. Derivation of the accelerated convergence formula *via* the Patterson function .. 454
3.4.9. Evaluation of the incomplete gamma function .. 454
3.4.10. Summation over the asymmetric unit and elimination of intramolecular energy terms .. 455
3.4.11. Reference formulae for particular values of n .. 455
3.4.12. Numerical illustrations .. 456

3.5. Extensions of the Ewald method for Coulomb interactions in crystals (T. A. Darden) .. 458
3.5.1. Introduction .. 458
3.5.2. Lattice sums of point charges .. 460
3.5.3. Generalization to Gaussian- and Hermite-based continuous charge distributions .. 471
3.5.4. Computational efficiency .. 474

PART 4. DIFFUSE SCATTERING AND RELATED TOPICS

4.1. Thermal diffuse scattering of X-rays and neutrons (B. T. M. Willis) .. 484
4.1.1. Introduction .. 484
4.1.2. Dynamics of three-dimensional crystals .. 484
4.1.3. Scattering of X-rays by thermal vibrations .. 487
4.1.4. Scattering of neutrons by thermal vibrations .. 488
4.1.5. Phonon dispersion relations .. 489
4.1.6. Measurement of elastic constants .. 490

4.2. Disorder diffuse scattering of X-rays and neutrons (F. Frey, H. Boysen and H. Jagodzinski) .. 492
4.2.1. Introduction .. 492
4.2.2. Basic scattering theory .. 493
4.2.3. Qualitative treatment of structural disorder .. 495
4.2.4. General guidelines for analysing a disorder problem .. 507
4.2.5. Quantitative interpretation .. 509
4.2.6. Disorder diffuse scattering from aperiodic crystals .. 526
4.2.7. Computer simulations and modelling .. 528
4.2.8. Experimental techniques and data evaluation .. 530

4.3. Diffuse scattering in electron diffraction (J. M. Cowley and J. K. Gjønnes) .. 540
4.3.1. Introduction .. 540
4.3.2. Inelastic scattering .. 541
4.3.3. Kinematical and pseudo-kinematical scattering .. 542
4.3.4. Dynamical scattering: Bragg scattering effects .. 542
4.3.5. Multislice calculations for diffraction and imaging .. 544
4.3.6. Qualitative interpretation of diffuse scattering of electrons .. 544

4.4. Scattering from mesomorphic structures (P. S. Pershan) .. 547
4.4.1. Introduction .. 547
4.4.2. The nematic phase .. 549
4.4.3. Smectic-A and smectic-C phases .. 551
4.4.4. Phases with in-plane order .. 554
4.4.5. Discotic phases .. 561
4.4.6. Other phases .. 561

4.5. Polymer crystallography (R. P. Millane and D. L. Dorset) .. 567
4.5.1. Overview (R. P. Millane and D. L. Dorset) .. 567
4.5.2. X-ray fibre diffraction analysis (R. P. Millane) .. 568
4.5.3. Electron crystallography of polymers (D. L. Dorset) .. 583

4.6. Reciprocal-space images of aperiodic crystals (W. Steurer and T. Haibach) .. 590
4.6.1. Introduction .. 590
4.6.2. The *n*-dimensional description of aperiodic crystals .. 591
4.6.3. Reciprocal-space images .. 598
4.6.4. Experimental aspects of the reciprocal-space analysis of aperiodic crystals .. 621

PART 5. DYNAMICAL THEORY AND ITS APPLICATIONS

5.1. Dynamical theory of X-ray diffraction (A. Authier) .. 626
5.1.1. Introduction .. 626
5.1.2. Fundamentals of plane-wave dynamical theory .. 626
5.1.3. Solutions of plane-wave dynamical theory .. 630
5.1.4. Standing waves .. 633
5.1.5. Anomalous absorption .. 633

5.1.6. Intensities of plane waves in transmission geometry .. 634
5.1.7. Intensity of plane waves in reflection geometry .. 638
5.1.8. Real waves .. 640
Appendix A5.1.1. Basic equations .. 642

5.2. Dynamical theory of electron diffraction (A. F. Moodie, J. M. Cowley and P. Goodman) .. 647
5.2.1. Introduction .. 647
5.2.2. The defining equations .. 647
5.2.3. Forward scattering .. 647
5.2.4. Evolution operator .. 648
5.2.5. Projection approximation – real-space solution .. 648
5.2.6. Semi-reciprocal space .. 648
5.2.7. Two-beam approximation .. 649
5.2.8. Eigenvalue approach .. 649
5.2.9. Translational invariance .. 650
5.2.10. Bloch-wave formulations .. 650
5.2.11. Dispersion surfaces .. 650
5.2.12. Multislice .. 651
5.2.13. Born series .. 651
5.2.14. Approximations .. 652

5.3. Dynamical theory of neutron diffraction (M. Schlenker and J.-P. Guigay) .. 654
5.3.1. Introduction .. 654
5.3.2. Comparison between X-rays and neutrons with spin neglected .. 654
5.3.3. Neutron spin, and diffraction by perfect magnetic crystals .. 655
5.3.4. Extinction in neutron diffraction (nonmagnetic case) .. 658
5.3.5. Effect of external fields on neutron scattering by perfect crystals .. 659
5.3.6. Experimental tests of the dynamical theory of neutron scattering .. 659
5.3.7. Applications of the dynamical theory of neutron scattering .. 660

Author index .. 665

Subject index .. 675

Preface

By Uri Shmueli

The purpose of Volume B of *International Tables for Crystallography* is to provide the user or reader with accounts of some well established topics, of importance to the science of crystallography, which are related in one way or another to the concepts of reciprocal lattice and, more generally, reciprocal space. Efforts have been made to extend the treatment of the various topics to include X-ray, electron, and neutron diffraction techniques, and thereby do some justice to the inclusion of the present Volume in the new series of *International Tables for Crystallography*.

An important crystallographic aspect of symmetry in reciprocal space, space-group-dependent expressions of trigonometric structure factors, already appears in Volume I of *International Tables for X-ray Crystallography*, and preliminary plans for incorporating this and other crystallographic aspects of reciprocal space in the new edition of *International Tables* date back to 1972. However, work on a volume of *International Tables for Crystallography*, largely dedicated to the subject of reciprocal space, began over ten years later. The present structure of Volume B, as determined in the years preceding the 1984 Hamburg congress of the International Union of Crystallography (IUCr), is due to (i) computer-controlled production of concise structure-factor tables, (ii) the ability to introduce many more aspects of reciprocal space – as a result of reducing the effort of producing the above tables, as well as their volume, and (iii) suggestions by the National Committees and individual crystallographers of some additional interesting topics. It should be pointed out that the initial plans for the present Volume and Volume C (*Mathematical, Physical and Chemical Tables*, edited by Professor A. J. C. Wilson), were formulated and approved during the same period.

The obviously delayed publication of Volume B is due to several reasons. Some minor delays were caused by a requirement that potential contributors should be approved by the Executive Committee prior to issuing relevant invitations. Much more serious delays were caused by authors who failed to deliver their contributions. In fact, some invited contributions had to be excluded from this first edition of Volume B. Some of the topics here treated are greatly extended, considerably updated or modern versions of similar topics previously treated in the old Volumes I, II, and IV. Most of the subjects treated in Volume B are new to *International Tables*.

I gratefully thank Professor A. J. C. Wilson, for suggesting that I edit this Volume and for sharing with me his rich editorial experience. I am indebted to those authors of Volume B who took my requests and deadlines seriously, and to the Computing Center of Tel Aviv University for computing facilities and time. Special thanks are due to Mrs Z. Stein (Tel Aviv University) for skilful assistance in numeric and symbolic programming, involved in my contributions to this Volume.

I am most grateful to many colleagues–crystallographers for encouragement, advice, and suggestions. In particular, thanks are due to Professors J. M. Cowley, P. Goodman and C. J. Humphreys, who served as Chairmen of the Commission on Electron Diffraction during the preparation of this Volume, for prompt and expert help at all stages of the editing. The kind assistance of Dr J. N. King, the Executive Secretary of the IUCr, is also gratefully acknowledged. Last, but certainly not least, I wish to thank Mr M. H. Dacombe, the Technical Editor of the IUCr, and his staff for the skilful and competent treatment of the variety of drafts and proofs out of which this Volume arose.

Preface to the second edition

By Uri Shmueli

The first edition of Volume B appeared in 1993, and was followed by a corrected reprint in 1996. Although practically all the material for the second edition was available in early 1997, its publication was delayed by the decision to translate all of Volume B, and indeed all the other volumes of *International Tables for Crystallography*, to Standard Generalized Markup Language (SGML) and thus make them available also in an electronic form suitable for modern publishing procedures.

During the preparation of the second edition, most chapters that appeared in the first edition have been corrected and/or revised, some were rather extensively updated, and five new chapters were added. The overall structure of the second edition is outlined below.

After an introductory chapter, Part 1 presents the reader with an account of structure-factor formalisms, an extensive treatment of the theory, algorithms and crystallographic applications of Fourier methods, and treatments of symmetry in reciprocal space. These are here enriched with more advanced aspects of representations of space groups in reciprocal space.

In Part 2, these general accounts are followed by detailed expositions of crystallographic statistics, the theory of direct methods, Patterson techniques, isomorphous replacement and anomalous scattering, and treatments of the role of electron microscopy and diffraction in crystal structure determination. The latter topic is here enhanced by applications of direct methods to electron crystallography.

Part 3, *Dual Bases in Crystallographic Computing*, deals with applications of reciprocal space to molecular geometry and 'best'-plane calculations, and contains a treatment of the principles of molecular graphics and modelling and their applications; it concludes with the presentation of a convergence-acceleration method, of importance in the computation of approximate lattice sums.

Part 4 contains treatments of various diffuse-scattering phenomena arising from crystal dynamics, disorder and low dimensionality (liquid crystals), and an exposition of the underlying theories and/or experimental evidence. The new additions to this part are treatments of polymer crystallography and of reciprocal-space images of aperiodic crystals.

Part 5 contains introductory treatments of the theory of the interaction of radiation with matter, the so-called dynamical theory, as applied to X-ray, electron and neutron diffraction techniques. The chapter on the dynamical theory of neutron diffraction is new.

I am deeply grateful to the authors of the new contributions for making their expertise available to Volume B and for their excellent collaboration. I also take special pleasure in thanking

those authors of the first edition who revised and updated their contributions in view of recent developments. Last but not least, I wish to thank all the authors for their contributions and their patience, and am grateful to those authors who took my requests seriously. I hope that the updating and revision of future editions will be much easier and more expedient, mainly because of the new format of *International Tables*.

Four friends and greatly respected colleagues who contributed to the second edition of Volume B are no longer with us. These are Professors Arthur J. C. Wilson, Peter Goodman, Verner Schomaker and Boris K. Vainshtein. I asked Professors Michiyoshi Tanaka, John Cowley and Douglas Dorset if they were prepared to answer queries related to the contributions of the late Peter Goodman and Boris K. Vainshtein to Chapter 2.5. I am most grateful for their prompt agreement.

This editorial work was carried out at the School of Chemistry and the Computing Center of Tel Aviv University. The facilities they put at my disposal are gratefully acknowledged on my behalf and on behalf of the IUCr. I wish to thank many colleagues for interesting conversations and advice, and in particular Professor Theo Hahn with whom I discussed at length problems regarding Volume B and *International Tables* in general.

Given all these expert contributions, the publication of this volume would not have been possible without the expertise and devotion of the Technical Editors of the IUCr. My thanks go to Mrs Sue King, for her cooperation during the early stages of the work on the second edition of Volume B, while the material was being collected, and to Dr Nicola Ashcroft, for her collaboration during the final stages of the production of the volume, for her most careful and competent treatment of the proofs, and last but not least for her tactful and friendly attitude.

Preface to the third edition

By Uri Shmueli

The second edition of Volume B appeared in 2001. Plans for the third edition included the addition of new chapters and sections, the substantial revision of several chapters that existed in the second edition and minor revisions and updating of existing chapters. The overall structure of Volume B remained unchanged.

In Part 1, Chapter 1.5 on classifications of space-group representations in reciprocal space has been extensively revised.

In Part 2, Chapter 2.2 on direct methods has been considerably extended to include applications of these methods to macromolecular crystallography. Chapter 2.3 on Patterson and molecular replacement techniques has been updated and extended. Section 2.5.3 on convergent-beam electron diffraction has been completely rewritten by a newly invited author, and Section 2.5.6 on three-dimensional reconstruction has been updated and extended by a newly invited author, who has also added Section 2.5.7 on single-particle reconstruction. The Foreword to Chapter 2.5 on electron diffraction and microscopy has also been revised.

In Part 3, Chapter 3.3 on computer graphics and molecular modelling has been enriched by Section 3.3.4 on the implementation of molecular graphics to small and medium-sized molecules, and a comprehensive Chapter 3.5 on modern extensions of Ewald methods has been added, dealing with (i) inclusion of fast Fourier transforms in the computation of sums and (ii) departure from the point-charge model in Ewald summations.

In Part 4, Chapter 4.1 on thermal diffuse scattering of X-rays and neutrons has been updated, and Chapter 4.2 on disorder diffuse scattering of X-rays and neutrons has been extensively revised and updated.

Minor updates and corrections have also been made to several existing chapters and sections in all the parts of the volume.

My gratitude is extended to the authors of new contributions and to the authors of the first and second editions of the volume for significant revisions of their chapters and sections in view of new developments. I wish to thank all the authors for their excellent collaboration and for sharing with the *International Tables for Crystallography* their expertise. I hope that the tradition of keeping the contributions up to date will also persist in future editions of Volume B. This will be aided by significant improvements in various aspects of technical editing which were already apparent in the preparation of this edition.

Three greatly respected friends and colleagues who contributed to this and previous editions of Volume B passed away after the second edition of Volume B was published. These are Professors John Cowley, Boris Zvyagin and Donald Williams. I asked Professors John Spence, Douglas Dorset and Pawel Penczek to take care of any questions about the articles of the late John Cowley, Boris Zvyagin and Boris Vainshtein in Chapter 2.5, and Dr Bill Smith to answer any questions about Chapter 3.4 by the late Donald Williams. They all agreed promptly and I am most grateful for this.

My editorial work was carried out at the School of Chemistry of Tel Aviv University and I wish to acknowledge gratefully the facilities that were put at my disposal. I am grateful to many friends and colleagues for interesting conversations and exchanges related to this volume. Thanks are also due to my friends from the IUCr office in Chester for their helpful interest.

Finally, I think that the publication of the third edition of Volume B would not have been possible without the competent, tactful and friendly collaboration of Dr Nicola Ashcroft, the Technical Editor of this project during all the stages of the preparation of this edition.

International Tables for Crystallography (2010). Vol. B, Chapter 1.1, pp. 2–9.

1.1. Reciprocal space in crystallography

By U. Shmueli

1.1.1. Introduction

The purpose of this chapter is to provide an introduction to several aspects of reciprocal space, which are of general importance in crystallography and which appear in the various chapters and sections to follow. We first summarize the basic definitions and briefly inspect some fundamental aspects of crystallography, while recalling that they can be usefully and simply discussed in terms of the concept of the reciprocal lattice. This introductory section is followed by a summary of the basic relationships between the direct and associated reciprocal lattices. We then introduce the elements of tensor-algebraic formulation of such dual relationships, with emphasis on those that are important in many applications of reciprocal space to crystallographic algorithms. We proceed with a section that demonstrates the role of mutually reciprocal bases in transformations of coordinates and conclude with a brief outline of some important analytical aspects of reciprocal space, most of which are further developed in other parts of this volume.

1.1.2. Reciprocal lattice in crystallography

The notion of mutually reciprocal triads of vectors dates back to the introduction of vector calculus by J. Willard Gibbs in the 1880s (*e.g.* Wilson, 1901). This concept appeared to be useful in the early interpretations of diffraction from single crystals (Ewald, 1913; Laue, 1914) and its first detailed exposition and the recognition of its importance in crystallography can be found in Ewald's (1921) article. The following free translation of Ewald's (1921) introduction, presented in a somewhat different notation, may serve the purpose of this section:

> To the set of $\mathbf{a}_i$, there corresponds in the vector calculus a set of 'reciprocal vectors' $\mathbf{b}_i$, which are defined (by Gibbs) by the following properties:
>
> $$\mathbf{a}_i \cdot \mathbf{b}_k = 0 \quad (\text{for } i \neq k) \tag{1.1.2.1}$$
>
> and
>
> $$\mathbf{a}_i \cdot \mathbf{b}_i = 1, \tag{1.1.2.2}$$
>
> where i and k may each equal 1, 2 or 3. The first equation, (1.1.2.1), says that each vector $\mathbf{b}_k$ is perpendicular to two vectors $\mathbf{a}_i$, as follows from the vanishing scalar products. Equation (1.1.2.2) provides the norm of the vector $\mathbf{b}_i$: the length of this vector must be chosen such that the projection of $\mathbf{b}_i$ on the direction of $\mathbf{a}_i$ has the length $1/a_i$, where a_i is the magnitude of the vector $\mathbf{a}_i$ ....

The consequences of equations (1.1.2.1) and (1.1.2.2) were elaborated by Ewald (1921) and are very well documented in the subsequent literature, crystallographic as well as other.

As is well known, the reciprocal lattice occupies a rather prominent position in crystallography and there are nearly as many accounts of its importance as there are crystallographic texts. It is not intended to review its applications, in any detail, in the present section; this is done in the remaining chapters and sections of the present volume. It seems desirable, however, to mention by way of an introduction some fundamental geometrical, physical and mathematical aspects of crystallography, and try to give a unified demonstration of the usefulness of mutually reciprocal bases as an interpretive tool.

Let us assume that the coordinates of all the (direct) lattice points are integers. This can only be true for *P*-type lattices. Consider the equation of a lattice plane in the direct lattice. It can be shown (*e.g.* Buerger, 1941; also Shmueli, 2007) that this equation is given by

$$hx + ky + lz = n, \tag{1.1.2.3}$$

where h, k and l, known as Miller indices of the (hkl) lattice plane, are (under the above assumption) relatively prime integers (*i.e.* do not have a common factor other than $+1$ or -1). In this equation, x, y and z are the coordinates of any point lying in the plane and are expressed as fractions of the magnitudes of the basis vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ of the direct lattice, and n is an integer denoting the serial number of the lattice plane within the family of parallel and equidistant (hkl) planes. The interplanar spacing is denoted by d_{hkl}, the value $n = 0$ corresponding to the (hkl) plane passing through the origin.

Let $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ and $\mathbf{r}_\mathrm{L} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$, where u, v, w are any integers, denote the position vectors of the point xyz and a lattice point uvw lying in the plane (1.1.2.3), respectively, and assume that $\mathbf{r}$ and $\mathbf{r}_\mathrm{L}$ are different vectors. If the plane normal is denoted by $\mathbf{N}$, where $\mathbf{N}$ is proportional to the vector product of two in-plane lattice vectors, the vector form of the equation of the lattice plane becomes

$$\mathbf{N} \cdot (\mathbf{r} - \mathbf{r}_\mathrm{L}) = 0 \quad \text{or} \quad \mathbf{N} \cdot \mathbf{r} = \mathbf{N} \cdot \mathbf{r}_\mathrm{L}. \tag{1.1.2.4}$$

For equations (1.1.2.3) and (1.1.2.4) to be identical, the plane normal $\mathbf{N}$ must satisfy the requirement that $\mathbf{N} \cdot \mathbf{r}_\mathrm{L} = n$, where n is an (unrestricted) integer.

While the Miller indices of lattice planes in *P*-type lattices must be relatively prime, if the direct lattice is based on a non-primitive unit cell (any centring type) the Miller indices of some lattice planes are no longer relatively prime (*e.g.* Nespolo, 2015).

Let us now consider the basic diffraction relations (*e.g.* Lipson & Cochran, 1966). Suppose a parallel beam of monochromatic radiation, of wavelength λ, falls on a lattice of identical point scatterers. If it is assumed that the scattering is elastic, *i.e.* there is no change of the wavelength during this process, the wavevectors of the incident and scattered radiation have the same magnitude, which can conveniently be taken as $1/\lambda$. A consideration of path and phase differences between the waves outgoing from two point scatterers separated by the lattice vector $\mathbf{r}_\mathrm{L}$ (defined as above) shows that the condition for their maximum constructive interference is given by

$$(\mathbf{s} - \mathbf{s}_0) \cdot \mathbf{r}_\mathrm{L} = n, \tag{1.1.2.5}$$

where $\mathbf{s}_0$ and $\mathbf{s}$ are the wavevectors of the incident and scattered beams, respectively, and n is an arbitrary integer.

Since $\mathbf{r}_\mathrm{L} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$, where u, v and w are unrestricted integers, equation (1.1.2.5) is equivalent to the equations of Laue:

$$\mathbf{h} \cdot \mathbf{a} = h, \quad \mathbf{h} \cdot \mathbf{b} = k, \quad \mathbf{h} \cdot \mathbf{c} = l, \tag{1.1.2.6}$$

where $\mathbf{h} = \mathbf{s} - \mathbf{s}_0$ is the diffraction vector, and h, k and l are integers corresponding to orders of diffraction from the three-dimensional lattice (Lipson & Cochran, 1966). The diffraction vector thus has to satisfy a condition that is analogous to that imposed on the normal to a lattice plane.

The next relevant aspect to be commented on is the Fourier expansion of a function having the periodicity of the crystal lattice. Such functions are *e.g.* the electron density, the density of nuclear matter and the electrostatic potential in the crystal, which are the operative definitions of crystal structure in X-ray, neutron and electron-diffraction methods of crystal structure determination. A Fourier expansion of such a periodic function may be

Copyright © 2010 International Union of Crystallography

thought of as a superposition of waves (*e.g.* Buerger, 1959), with wavevectors related to the interplanar spacings d_{hkl}, in the crystal lattice. Denoting the wavevector of a Fourier wave by $\mathbf{g}$ (a function of hkl), the phase of the Fourier wave at the point $\mathbf{r}$ in the crystal is given by $2\pi\mathbf{g}\cdot\mathbf{r}$, and the triple Fourier series corresponding to the expansion of the periodic function, say $G(\mathbf{r})$, can be written as

$$G(\mathbf{r}) = \sum_{\mathbf{g}} C(\mathbf{g}) \exp(-2\pi i \mathbf{g}\cdot\mathbf{r}), \tag{1.1.2.7}$$

where $C(\mathbf{g})$ are the amplitudes of the Fourier waves, or Fourier coefficients, which are related to the experimental data. Numerous examples of such expansions appear throughout this volume.

The permissible wavevectors in the above expansion are restricted by the periodicity of the function $G(\mathbf{r})$. Since, by definition, $G(\mathbf{r}) = G(\mathbf{r}+\mathbf{r}_{\rm L})$, where $\mathbf{r}_{\rm L}$ is a direct-lattice vector, the right-hand side of (1.1.2.7) must remain unchanged when $\mathbf{r}$ is replaced by $\mathbf{r}+\mathbf{r}_{\rm L}$. This, however, can be true only if the scalar product $\mathbf{g}\cdot\mathbf{r}_{\rm L}$ is an integer.

Each of the above three aspects of crystallography may lead, independently, to a useful introduction of the reciprocal vectors, and there are many examples of this in the literature. It is interesting, however, to consider the representation of the equation

$$\mathbf{v}\cdot\mathbf{r}_{\rm L} = n, \tag{1.1.2.8}$$

which is common to all three, in its most convenient form. Obviously, the vector $\mathbf{v}$ which stands for the plane normal, the diffraction vector, and the wavevector in a Fourier expansion, may still be referred to any permissible basis and so may $\mathbf{r}_{\rm L}$, by an appropriate transformation.

Let $\mathbf{v} = U\mathbf{A} + V\mathbf{B} + W\mathbf{C}$, where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are linearly independent vectors. Equation (1.1.2.8) can then be written as

$$(U\mathbf{A} + V\mathbf{B} + W\mathbf{C})\cdot(u\mathbf{a} + v\mathbf{b} + w\mathbf{c}) = n, \tag{1.1.2.9}$$

or, in matrix notation,

$$(UVW)\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\\ \mathbf{C}\end{pmatrix}\cdot(\mathbf{abc})\begin{pmatrix}u\\ v\\ w\end{pmatrix} = n, \tag{1.1.2.10}$$

or

$$(UVW)\begin{pmatrix}\mathbf{A}\cdot\mathbf{a} & \mathbf{A}\cdot\mathbf{b} & \mathbf{A}\cdot\mathbf{c}\\ \mathbf{B}\cdot\mathbf{a} & \mathbf{B}\cdot\mathbf{b} & \mathbf{B}\cdot\mathbf{c}\\ \mathbf{C}\cdot\mathbf{a} & \mathbf{C}\cdot\mathbf{b} & \mathbf{C}\cdot\mathbf{c}\end{pmatrix}\begin{pmatrix}u\\ v\\ w\end{pmatrix} = n. \tag{1.1.2.11}$$

The simplest representation of equation (1.1.2.8) results when the matrix of scalar products in (1.1.2.11) reduces to a unit matrix. This can be achieved (i) by choosing the basis vectors **ABC** to be orthonormal to the basis vectors **abc**, while requiring that the components of $\mathbf{r}_{\rm L}$ be integers, or (ii) by requiring that the bases **ABC** and **abc** coincide with the same orthonormal basis, *i.e.* expressing both $\mathbf{v}$ and $\mathbf{r}_{\rm L}$, in (1.1.2.8), in the same Cartesian system. If we choose the first alternative, it is seen that:

(1) The components of the vector $\mathbf{v}$, and hence those of $\mathbf{N}$, $\mathbf{h}$ and $\mathbf{g}$, are of necessity integers, since u, v and w are already integral. The components of $\mathbf{v}$ include Miller indices in the case of the lattice plane, they coincide with the orders of diffraction from a three-dimensional lattice of scatterers, and correspond to the summation indices in the triple Fourier series (1.1.2.7).

(2) The basis vectors $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are reciprocal to $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, as can be seen by comparing the scalar products in (1.1.2.11) with those in (1.1.2.1) and (1.1.2.2). In fact, the bases **ABC** and **abc** are mutually reciprocal. Since there are no restrictions on the integers U, V and W, the vector $\mathbf{v}$ belongs to a lattice which, on account of its basis, is called the reciprocal lattice.

It follows that, at least in the present case, algebraic simplicity goes together with ease of interpretation, which certainly accounts for much of the importance of the reciprocal lattice in crystallography. The second alternative of reducing the matrix in (1.1.2.11) to a unit matrix, a transformation of (1.1.2.8) to a Cartesian system, leads to non-integral components of the vectors, which makes any interpretation of $\mathbf{v}$ or $\mathbf{r}_{\rm L}$ much less transparent. However, transformations to Cartesian systems are often very useful in crystallographic computing and will be discussed below (see also Chapters 2.3 and 3.3 in this volume).

We shall, in what follows, abandon all the temporary notation used above and write the reciprocal-lattice vector as

$$\mathbf{h} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{1.1.2.12}$$

or

$$\mathbf{h} = h_1\mathbf{a}^1 + h_2\mathbf{a}^2 + h_3\mathbf{a}^3 = \sum_{i=1}^{3} h_i\mathbf{a}^i, \tag{1.1.2.13}$$

and denote the direct-lattice vectors by $\mathbf{r}_{\rm L} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$, as above, or by

$$\mathbf{r}_{\rm L} = u^1\mathbf{a}_1 + u^2\mathbf{a}_2 + u^3\mathbf{a}_3 = \sum_{i=1}^{3} u^i\mathbf{a}_i. \tag{1.1.2.14}$$

The representations (1.1.2.13) and (1.1.2.14) are used in the tensor-algebraic formulation of the relationships between mutually reciprocal bases (see Section 1.1.4 below).

1.1.3. Fundamental relationships

We now present a brief derivation and a summary of the most important relationships between the direct and the reciprocal bases. The usual conventions of vector algebra are observed and the results are presented in the conventional crystallographic notation. Equations (1.1.2.1) and (1.1.2.2) now become

$$\mathbf{a}\cdot\mathbf{b}^* = \mathbf{a}\cdot\mathbf{c}^* = \mathbf{b}\cdot\mathbf{a}^* = \mathbf{b}\cdot\mathbf{c}^* = \mathbf{c}\cdot\mathbf{a}^* = \mathbf{c}\cdot\mathbf{b}^* = 0 \tag{1.1.3.1}$$

and

$$\mathbf{a}\cdot\mathbf{a}^* = \mathbf{b}\cdot\mathbf{b}^* = \mathbf{c}\cdot\mathbf{c}^* = 1, \tag{1.1.3.2}$$

respectively, and the relationships are obtained as follows.

1.1.3.1. Basis vectors

It is seen from (1.1.3.1) that $\mathbf{a}^*$ must be proportional to the vector product of $\mathbf{b}$ and $\mathbf{c}$,

$$\mathbf{a}^* = K(\mathbf{b}\times\mathbf{c}),$$

and, since $\mathbf{a}\cdot\mathbf{a}^* = 1$, the proportionality constant K equals $1/[\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})]$. The mixed product $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ can be interpreted as the positive volume of the unit cell in the direct lattice only if $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ form a *right-handed* set. If the above condition is fulfilled, we obtain

$$\mathbf{a}^* = \frac{\mathbf{b}\times\mathbf{c}}{V}, \quad \mathbf{b}^* = \frac{\mathbf{c}\times\mathbf{a}}{V}, \quad \mathbf{c}^* = \frac{\mathbf{a}\times\mathbf{b}}{V} \tag{1.1.3.3}$$

and analogously

$$\mathbf{a} = \frac{\mathbf{b}^*\times\mathbf{c}^*}{V^*}, \quad \mathbf{b} = \frac{\mathbf{c}^*\times\mathbf{a}^*}{V^*}, \quad \mathbf{c} = \frac{\mathbf{a}^*\times\mathbf{b}^*}{V^*}, \tag{1.1.3.4}$$

where V and V^* are the volumes of the unit cells in the associated direct and reciprocal lattices, respectively. Use has been made of the fact that the mixed product, say $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$, remains unchanged under cyclic rearrangement of the vectors that appear in it.

1.1.3.2. Volumes

The reciprocal relationship of V and V^* follows readily. We have from equations (1.1.3.2), (1.1.3.3) and (1.1.3.4)

$$\mathbf{c}\cdot\mathbf{c}^* = \frac{(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}^*\times\mathbf{b}^*)}{VV^*} = 1.$$

If we make use of the vector identity

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}), \quad (1.1.3.5)$$

and equations (1.1.3.1) and (1.1.3.2), it is seen that $V^* = 1/V$.

1.1.3.3. Angular relationships

The relationships of the angles α, β, γ between the pairs of vectors (**b**, **c**), (**c**, **a**) and (**a**, **b**), respectively, and the angles $\alpha^*, \beta^*, \gamma^*$ between the corresponding pairs of reciprocal basis vectors, can be obtained by simple vector algebra. For example, we have from (1.1.3.3):

(i) $\mathbf{b}^*\cdot\mathbf{c}^* = b^*c^*\cos\alpha^*$, with

$$b^* = \frac{ca\sin\beta}{V} \quad \text{and} \quad c^* = \frac{ab\sin\gamma}{V}$$

and (ii)

$$\mathbf{b}^*\cdot\mathbf{c}^* = \frac{(\mathbf{c}\times\mathbf{a})\cdot(\mathbf{a}\times\mathbf{b})}{V^2}.$$

If we make use of the identity (1.1.3.5), and compare the two expressions for $\mathbf{b}^*\cdot\mathbf{c}^*$, we readily obtain

$$\cos\alpha^* = \frac{\cos\beta\cos\gamma - \cos\alpha}{\sin\beta\sin\gamma}. \quad (1.1.3.6)$$

Similarly,

$$\cos\beta^* = \frac{\cos\gamma\cos\alpha - \cos\beta}{\sin\gamma\sin\alpha} \quad (1.1.3.7)$$

and

$$\cos\gamma^* = \frac{\cos\alpha\cos\beta - \cos\gamma}{\sin\alpha\sin\beta}. \quad (1.1.3.8)$$

The expressions for the cosines of the direct angles in terms of those of the reciprocal ones are analogous to (1.1.3.6)–(1.1.3.8). For example,

$$\cos\alpha = \frac{\cos\beta^*\cos\gamma^* - \cos\alpha^*}{\sin\beta^*\sin\gamma^*}.$$

1.1.3.4. Matrices of metric tensors

Various computational and algebraic aspects of mutually reciprocal bases are most conveniently expressed in terms of the metric tensors of these bases. The tensors will be treated in some detail in the next section, and only the definitions of their matrices are given and interpreted below.

Consider the length of the vector $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$. This is given by

$$|\mathbf{r}| = [(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})\cdot(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})]^{1/2} \quad (1.1.3.9)$$

and can be written in matrix form as

$$|\mathbf{r}| = [\boldsymbol{x}^T\boldsymbol{G}\boldsymbol{x}]^{1/2}, \quad (1.1.3.10)$$

where

$$\boldsymbol{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \boldsymbol{x}^T = (xyz)$$

and

$$\boldsymbol{G} = \begin{pmatrix} \mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c} \\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c} \\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c} \end{pmatrix} \quad (1.1.3.11)$$

$$= \begin{pmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ba\cos\gamma & b^2 & bc\cos\alpha \\ ca\cos\beta & cb\cos\alpha & c^2 \end{pmatrix}. \quad (1.1.3.12)$$

This is the matrix of the metric tensor of the direct basis, or briefly the direct metric. The corresponding reciprocal metric is given by

$$\boldsymbol{G}^* = \begin{pmatrix} \mathbf{a}^*\cdot\mathbf{a}^* & \mathbf{a}^*\cdot\mathbf{b}^* & \mathbf{a}^*\cdot\mathbf{c}^* \\ \mathbf{b}^*\cdot\mathbf{a}^* & \mathbf{b}^*\cdot\mathbf{b}^* & \mathbf{b}^*\cdot\mathbf{c}^* \\ \mathbf{c}^*\cdot\mathbf{a}^* & \mathbf{c}^*\cdot\mathbf{b}^* & \mathbf{c}^*\cdot\mathbf{c}^* \end{pmatrix} \quad (1.1.3.13)$$

$$= \begin{pmatrix} a^{*2} & a^*b^*\cos\gamma^* & a^*c^*\cos\beta^* \\ b^*a^*\cos\gamma^* & b^{*2} & b^*c^*\cos\alpha^* \\ c^*a^*\cos\beta^* & c^*b^*\cos\alpha^* & c^{*2} \end{pmatrix}. \quad (1.1.3.14)$$

The matrices $\boldsymbol{G}$ and $\boldsymbol{G}^*$ are of fundamental importance in crystallographic computations and transformations of basis vectors and coordinates from direct to reciprocal space and *vice versa.* Examples of applications are presented in Part 3 of this volume and in the remaining sections of this chapter.

It can be shown (*e.g.* Buerger, 1941) that the determinants of $\boldsymbol{G}$ and $\boldsymbol{G}^*$ equal the squared volumes of the direct and reciprocal unit cells, respectively. Thus,

$$\det(\boldsymbol{G}) = [\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})]^2 = V^2 \quad (1.1.3.15)$$

and

$$\det(\boldsymbol{G}^*) = [\mathbf{a}^*\cdot(\mathbf{b}^*\times\mathbf{c}^*)]^2 = V^{*2}, \quad (1.1.3.16)$$

and a direct expansion of the determinants, from (1.1.3.12) and (1.1.3.14), leads to

$$V = abc(1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma)^{1/2} \quad (1.1.3.17)$$

and

$$V^* = a^*b^*c^*(1 - \cos^2\alpha^* - \cos^2\beta^* - \cos^2\gamma^* + 2\cos\alpha^*\cos\beta^*\cos\gamma^*)^{1/2}. \quad (1.1.3.18)$$

The following algorithm has been found useful in computational applications of the above relationships to calculations in reciprocal space (*e.g.* data reduction) and in direct space (*e.g.* crystal geometry).

(1) Input the direct unit-cell parameters and construct the matrix of the metric tensor [*cf.* equation (1.1.3.12)].

(2) Compute the determinant of the matrix $\boldsymbol{G}$ and find the inverse matrix, $\boldsymbol{G}^{-1}$; this inverse matrix is just $\boldsymbol{G}^*$, the matrix of the metric tensor of the reciprocal basis (see also Section 1.1.4 below).

(3) Use the elements of $\boldsymbol{G}^*$, and equation (1.1.3.14), to obtain the parameters of the reciprocal unit cell.

The direct and reciprocal sets of unit-cell parameters, as well as the corresponding metric tensors, are now available for further calculations.

Explicit relations between direct- and reciprocal-lattice parameters, valid for the various crystal systems, are given in most

textbooks on crystallography [see also Chapters 1.1 and 1.2 of Volume C (Koch, 2004)].

1.1.4. Tensor-algebraic formulation

The present section summarizes the tensor-algebraic properties of mutually reciprocal sets of basis vectors, which are of importance in the various aspects of crystallography. This is not intended to be a systematic treatment of tensor algebra; for more thorough expositions of the subject the reader is referred to relevant crystallographic texts (*e.g.* Patterson, 1967; Sands, 1982), and other texts in the physical and mathematical literature that deal with tensor algebra and analysis.

Let us first recall that symbolic vector and matrix notations, in which basis vectors and coordinates do not appear explicitly, are often helpful in qualitative considerations. If, however, an expression has to be evaluated, the various quantities appearing in it must be presented in component form. One of the best ways to achieve a concise presentation of geometrical expressions in component form, while retaining much of their 'transparent' symbolic character, is their tensor-algebraic formulation.

1.1.4.1. Conventions

We shall adhere to the following conventions:

(i) Notation for direct and reciprocal basis vectors:

$$\mathbf{a} = \mathbf{a}_1, \mathbf{b} = \mathbf{a}_2, \mathbf{c} = \mathbf{a}_3$$
$$\mathbf{a}^* = \mathbf{a}^1, \mathbf{b}^* = \mathbf{a}^2, \mathbf{c}^* = \mathbf{a}^3.$$

Subscripted quantities are associated in tensor algebra with *covariant*, and superscripted with *contravariant* transformation properties. Thus the basis vectors of the direct lattice are represented as covariant quantities and those of the reciprocal lattice as contravariant ones.

(ii) Summation convention: if an index appears twice in an expression, once as subscript and once as superscript, a summation over this index is thereby implied and the summation sign is omitted. For example,

$$\sum_i \sum_j x^i T_{ij} x^j \text{ will be written } x^i T_{ij} x^j$$

since both i and j conform to the convention. Such repeating indices are often called *dummy* indices. The implied summation over repeating indices is also often used even when the indices are at the same level and the coordinate system is Cartesian; there is no distinction between contravariant and covariant quantities in Cartesian frames of reference (see Chapter 3.3).

(iii) Components (coordinates) of vectors referred to the covariant basis are written as contravariant quantities, and *vice versa.* For example,

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} = x^1\mathbf{a}_1 + x^2\mathbf{a}_2 + x^3\mathbf{a}_3 = x^i\mathbf{a}_i$$
$$\mathbf{h} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* = h_1\mathbf{a}^1 + h_2\mathbf{a}^2 + h_3\mathbf{a}^3 = h_i\mathbf{a}^i.$$

1.1.4.2. Transformations

A familiar concept but a fundamental one in tensor algebra is the transformation of coordinates. For example, suppose that an atomic position vector is referred to two unit-cell settings as follows:

$$\mathbf{r} = x^k\mathbf{a}_k \tag{1.1.4.1}$$

and

$$\mathbf{r} = x'^k\mathbf{a}'_k. \tag{1.1.4.2}$$

Let us multiply both sides of (1.1.4.1) and (1.1.4.2), on the right, by the vectors $\mathbf{a}^m$, m = 1, 2, or 3, *i.e.* by the reciprocal vectors to the basis $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$. We obtain from (1.1.4.1)

$$x^k\mathbf{a}_k \cdot \mathbf{a}^m = x^k\delta^m_k = x^m,$$

where δ^m_k is the Kronecker symbol which equals 1 when $k = m$ and equals zero if $k \neq m$, and by comparison with (1.1.4.2) we have

$$x^m = x'^k T^m_k, \tag{1.1.4.3}$$

where $T^m_k = \mathbf{a}'_k \cdot \mathbf{a}^m$ is an element of the required transformation matrix. Of course, the same transformation could have been written as

$$x^m = T^m_k x'^k, \tag{1.1.4.4}$$

where $T^m_k = \mathbf{a}^m \cdot \mathbf{a}'_k$.

A *tensor* is a quantity that transforms as the product of coordinates, and the *rank* of a tensor is the number of transformations involved (Patterson, 1967; Sands, 1982). *E.g.* the product of two coordinates, as in the above example, transforms from the $\mathbf{a}'$ basis to the $\mathbf{a}$ basis as

$$x^m x^n = T^m_p T^n_q x'^p x'^q; \tag{1.1.4.5}$$

the same transformation law applies to the components of a contravariant tensor of rank two, the components of which are referred to the primed basis and are to be transformed to the unprimed one:

$$Q^{mn} = T^m_p T^n_q Q'^{pq}. \tag{1.1.4.6}$$

1.1.4.3. Scalar products

The expression for the scalar product of two vectors, say $\mathbf{u}$ and $\mathbf{v}$, depends on the bases to which the vectors are referred. If we admit only the covariant and contravariant bases defined above, we have four possible types of expression:

(I) $\mathbf{u} = u^i\mathbf{a}_i, \mathbf{v} = v^i\mathbf{a}_i$

$$\mathbf{u} \cdot \mathbf{v} = u^i v^j(\mathbf{a}_i \cdot \mathbf{a}_j) \equiv u^i v^j g_{ij}, \tag{1.1.4.7}$$

(II) $\mathbf{u} = \mathbf{u}_i\mathbf{a}^i, \mathbf{v} = v_i\mathbf{a}^i$

$$\mathbf{u} \cdot \mathbf{v} = u_i v_j(\mathbf{a}^i \cdot \mathbf{a}^j) \equiv u_i v_j g^{ij}, \tag{1.1.4.8}$$

(III) $\mathbf{u} = u^i\mathbf{a}_i, \mathbf{v} = v_i\mathbf{a}^i$

$$\mathbf{u} \cdot \mathbf{v} = u^i v_j(\mathbf{a}_i \cdot \mathbf{a}^j) \equiv u^i v_j \delta^j_i = u^i v_i, \tag{1.1.4.9}$$

(IV) $\mathbf{u} = u_i\mathbf{a}^i, \mathbf{v} = v^i\mathbf{a}_i$

$$\mathbf{u} \cdot \mathbf{v} = u_i v^j(\mathbf{a}^i \cdot \mathbf{a}_j) \equiv u_i v^j \delta^i_j = u_i v^i. \tag{1.1.4.10}$$

(i) The sets of scalar products $g_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j$ (1.1.4.7) and $g^{ij} = \mathbf{a}^i \cdot \mathbf{a}^j$ (1.1.4.8) are known as the *metric* tensors of the covariant (direct) and contravariant (reciprocal) bases, respectively; the corresponding matrices are presented in conventional notation in equations (1.1.3.11) and (1.1.3.13). Numerous applications of these tensors to the computation of distances and angles in crystals are given in Chapter 3.1.

(ii) Equations (1.1.4.7) to (1.1.4.10) furnish the relationships between the covariant and contravariant *components* of the same vector. Thus, comparing (1.1.4.7) and (1.1.4.9), we have

$$v_i = v^j g_{ij}. \tag{1.1.4.11}$$

Similarly, using (1.1.4.8) and (1.1.4.10) we obtain the inverse relationship

$$v^i = v_j g^{ij}. \tag{1.1.4.12}$$

The corresponding relationships between covariant and contravariant *bases* can now be obtained if we refer a vector, say **v**, to each of the bases

$$\mathbf{v} = v^i \mathbf{a}_i = v_k \mathbf{a}^k,$$

and make use of (1.1.4.11) and (1.1.4.12). Thus, *e.g.*,

$$v^i \mathbf{a}_i = (v_k g^{ik}) \mathbf{a}_i = v_k \mathbf{a}^k.$$

Hence

$$\mathbf{a}^k = g^{ik} \mathbf{a}_i \tag{1.1.4.13}$$

and, similarly,

$$\mathbf{a}_k = g_{ik} \mathbf{a}^i. \tag{1.1.4.14}$$

(iii) The tensors g_{ij} and g^{ij} are symmetric, by definition.

(iv) It follows from (1.1.4.11) and (1.1.4.12) or (1.1.4.13) and (1.1.4.14) that the matrices of the direct and reciprocal metric tensors are mutually inverse, *i.e.*

$$\begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix}^{-1} = \begin{pmatrix} g^{11} & g^{12} & g^{13} \\ g^{21} & g^{22} & g^{23} \\ g^{31} & g^{32} & g^{33} \end{pmatrix}, \tag{1.1.4.15}$$

and their determinants are mutually reciprocal.

1.1.4.4. Examples

There are numerous applications of tensor notation in crystallographic calculations, and many of them appear in the various chapters of this volume. We shall therefore present only a few examples.

(i) The (squared) magnitude of the diffraction vector $\mathbf{h} = h_i \mathbf{a}^i$ is given by

$$|h|^2 = \frac{4 \sin^2 \theta}{\lambda^2} = h_i h_j g^{ij}. \tag{1.1.4.16}$$

This concise relationship is a starting point in a derivation of unit-cell parameters from experimental data.

(ii) The structure factor, including explicitly anisotropic displacement tensors, can be written in symbolic matrix notation as

$$F(\boldsymbol{h}) = \sum_{j=1}^{N} f_{(j)} \exp(-\boldsymbol{h}^T \boldsymbol{\beta}_{(j)} \boldsymbol{h}) \exp(2\pi i \boldsymbol{h}^T \boldsymbol{r}_{(j)}), \tag{1.1.4.17}$$

where $\boldsymbol{\beta}_{(j)}$ is the matrix of the anisotropic displacement tensor of the jth atom. In tensor notation, with the quantities referred to their natural bases, the structure factor can be written as

$$F(h_1 h_2 h_3) = \sum_{j=1}^{N} f_{(j)} \exp(-h_i h_k \beta^{ik}_{(j)}) \exp(2\pi i h_i x^i_{(j)}), \tag{1.1.4.18}$$

and similarly concise expressions can be written for the derivatives of the structure factor with respect to the positional and displacement parameters. The summation convention applies only to indices denoting components of vectors and tensors; the atom subscript j in (1.1.4.18) clearly does not qualify, and to indicate this it has been surrounded by parentheses.

(iii) Geometrical calculations, such as those described in the chapters of Part 3, may be carried out in any convenient basis but there are often some definite advantages to computations that are referred to the natural, non-Cartesian bases (see Chapter 3.1). Usually, the output positional parameters from structure refinement are available as contravariant components of the atomic position vectors. If we transform them by (1.1.4.11) to their covariant form, and store these covariant components of the atomic position vectors, the computation of scalar products using equations (1.1.4.9) or (1.1.4.10) is *almost as efficient as it would be if the coordinates were referred to a Cartesian system.* For example, the right-hand side of the vector identity (1.1.3.5), which is employed in the computation of dihedral angles, can be written as

$$(A_i C^i)(B_j D^j) - (A_k D^k)(B_l C^l).$$

This is a typical application of reciprocal space to ordinary direct-space computations.

(iv) We wish to derive a tensor formulation of the vector product, along similar lines to those of Chapter 3.1. As with the scalar product, there are several such formulations and we choose that which has both vectors, say **u** and **v**, and the resulting product, $\mathbf{u} \times \mathbf{v}$, referred to a covariant basis. We have

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= u^i \mathbf{a}_i \times v^j \mathbf{a}_j \\ &= u^i v^j (\mathbf{a}_i \times \mathbf{a}_j). \end{aligned} \tag{1.1.4.19}$$

If we make use of the relationships (1.1.3.3) between the direct and reciprocal basis vectors, it can be verified that

$$\mathbf{a}_i \times \mathbf{a}_j = V e_{kij} \mathbf{a}^k, \tag{1.1.4.20}$$

where V is the volume of the unit cell and the antisymmetric tensor e_{kij} equals $+1$, -1, or 0 according as kij is an even permutation of 123, an odd permutation of 123 or any two of the indices kij have the same value, respectively. We thus have

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= V e_{kij} u^i v^j \mathbf{a}^k \\ &= V g^{lk} e_{kij} u^i v^j \mathbf{a}_l, \end{aligned} \tag{1.1.4.21}$$

since by (1.1.4.13), $\mathbf{a}^k = g^{lk} \mathbf{a}_l$.

(v) *The rotation operator.* The general formulation of an expression for the rotation operator is of interest in crystal structure determination by Patterson techniques (see Chapter 2.3) and in molecular modelling (see Chapter 3.3), and another well known crystallographic application of this device is the derivation of the translation, libration and screw-motion tensors by the method of Schomaker & Trueblood (1968), discussed in Part 8 of Volume C (*IT* C, 2004) and in Chapter 1.2 of this volume. A digression on an elementary derivation of the above seems to be worthwhile.

Suppose we wish to rotate the vector **r**, about an axis coinciding with the unit vector **k**, through the angle θ and in the positive sense, *i.e.* an observer looking in the direction of $+\mathbf{k}$ will see **r** rotating in the clockwise sense. The vectors **r**, **k** and the rotated (target) vector $\mathbf{r}'$ are referred to an origin on the axis of rotation (see Fig. 1.1.4.1). Our purpose is to express $\mathbf{r}'$ in terms of **r**, **k** and θ by a general vector formula, and represent the components of the rotated vectors in coordinate systems that might be of interest.

Let us decompose the vector **r** and the (target) vector $\mathbf{r}'$ into their components which are parallel ($\parallel$) and perpendicular ($\perp$) to the axis of rotation:

$$\mathbf{r} = \mathbf{r}_\parallel + \mathbf{r}_\perp \tag{1.1.4.22}$$

and

$$\mathbf{r}' = \mathbf{r}'_\parallel + \mathbf{r}'_\perp. \tag{1.1.4.23}$$

It can be seen from Fig. 1.1.4.1 that the parallel components of **r** and $\mathbf{r}'$ are

$$\mathbf{r}_\parallel = \mathbf{r}'_\parallel = \mathbf{k}(\mathbf{k} \cdot \mathbf{r}) \tag{1.1.4.24}$$

and thus

$$\mathbf{r}_\perp = \mathbf{r} - \mathbf{k}(\mathbf{k} \cdot \mathbf{r}). \tag{1.1.4.25}$$

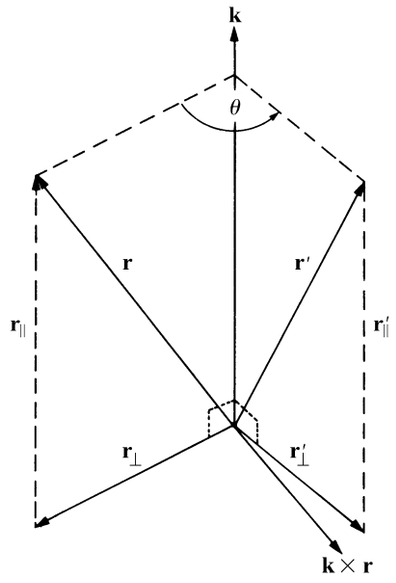


Fig. 1.1.4.1. Derivation of the general expression for the rotation operator. The figure illustrates schematically the decompositions and other simple geometrical considerations required for the derivation outlined in equations (1.1.4.22)–(1.1.4.28).

Only a suitable expression for $\mathbf{r}'_\perp$ is missing. We can find this by decomposing $\mathbf{r}'_\perp$ into its components (i) parallel to $\mathbf{r}_\perp$ and (ii) parallel to $\mathbf{k}\times\mathbf{r}_\perp$. We have, as in (1.1.4.24),

$$\mathbf{r}'_\perp = \frac{\mathbf{r}_\perp}{|\mathbf{r}_\perp|}\left(\frac{\mathbf{r}_\perp}{|\mathbf{r}_\perp|}\cdot\mathbf{r}'_\perp\right) + \frac{\mathbf{k}\times\mathbf{r}_\perp}{|\mathbf{k}\times\mathbf{r}_\perp|}\left(\frac{\mathbf{k}\times\mathbf{r}_\perp}{|\mathbf{k}\times\mathbf{r}_\perp|}\cdot\mathbf{r}'_\perp\right). \quad (1.1.4.26)$$

We observe, using Fig. 1.1.4.1, that

$$|\mathbf{r}'_\perp| = |\mathbf{r}_\perp| = |\mathbf{k}\times\mathbf{r}_\perp|$$

and

$$\mathbf{k}\times\mathbf{r}_\perp = \mathbf{k}\times\mathbf{r},$$

and, further,

$$\mathbf{r}'_\perp\cdot\mathbf{r}_\perp = |\mathbf{r}_\perp|^2\cos\theta$$

and

$$\mathbf{r}'_\perp\cdot(\mathbf{k}\times\mathbf{r}_\perp) = \mathbf{k}\cdot(\mathbf{r}'_\perp\times\mathbf{r}_\perp) = |\mathbf{r}_\perp|^2\sin\theta,$$

since the unit vector $\mathbf{k}$ is perpendicular to the plane containing the vectors $\mathbf{r}_\perp$ and $\mathbf{r}'_\perp$. Equation (1.1.4.26) now reduces to

$$\mathbf{r}'_\perp = \mathbf{r}_\perp\cos\theta + (\mathbf{k}\times\mathbf{r})\sin\theta \quad (1.1.4.27)$$

and equations (1.1.4.23), (1.1.4.25) and (1.1.4.27) lead to the required result

$$\mathbf{r}' = \mathbf{k}(\mathbf{k}\cdot\mathbf{r})(1-\cos\theta) + \mathbf{r}\cos\theta + (\mathbf{k}\times\mathbf{r})\sin\theta. \quad (1.1.4.28)$$

The above general expression can be written as a linear transformation by referring the vectors to an appropriate basis or bases. We choose here $\mathbf{r} = x^j\mathbf{a}_j$, $\mathbf{r}' = x'^i\mathbf{a}_i$ and assume that the components of $\mathbf{k}$ are available in the direct and reciprocal bases.

If we make use of equations (1.1.4.9) and (1.1.4.21), (1.1.4.28) can be written as

$$x'^i = k^i(k_j x^j)(1-\cos\theta) + \delta^i_j x^j\cos\theta + Vg^{im}e_{mpj}k^p x^j\sin\theta, \quad (1.1.4.29)$$

or briefly

$$x'^i = R^i_j x^j, \quad (1.1.4.30)$$

where

$$R^i_j = k^i k_j(1-\cos\theta) + \delta^i_j\cos\theta + Vg^{im}e_{mpj}k^p\sin\theta \quad (1.1.4.31)$$

is a matrix element of the rotation operator $\mathbf{R}$ which carries the vector $\mathbf{r}$ into the vector $\mathbf{r}'$. Of course, the representation (1.1.4.31) of $\mathbf{R}$ depends on our choice of reference bases.

If all the vectors are referred to a Cartesian basis, that is three orthogonal unit vectors, the direct and reciprocal metric tensors reduce to a unit tensor, there is no difference between covariant and contravariant quantities, and equation (1.1.4.31) reduces to

$$R_{ij} = k_i k_j(1-\cos\theta) + \delta_{ij}\cos\theta + e_{ipj}k_p\sin\theta, \quad (1.1.4.32)$$

where all the indices have been taken as subscripts, but the summation convention is still observed. The relative simplicity of (1.1.4.32), as compared to (1.1.4.31), often justifies the transformation of all the vector quantities to a Cartesian basis. This is certainly the case for any extensive calculation in which covariances of the structural parameters are not considered.

1.1.5. Transformations

1.1.5.1. Transformations of coordinates

It happens rather frequently that a vector referred to a given basis has to be re-expressed in terms of another basis, and it is then required to find the relationship between the components (coordinates) of the vector in the two bases. Such situations have already been indicated in the previous section. The purpose of the present section is to give a general method of finding such relationships (transformations), and discuss some simplifications brought about by the use of mutually reciprocal and Cartesian bases. We do not assume anything about the bases, in the general treatment, and hence the tensor formulation of Section 1.1.4 is not appropriate at this stage.

Let

$$\mathbf{r} = \sum_{j=1}^{3} u_j(1)\mathbf{c}_j(1) \quad (1.1.5.1)$$

and

$$\mathbf{r} = \sum_{j=1}^{3} u_j(2)\mathbf{c}_j(2) \quad (1.1.5.2)$$

be the *given* and *required* representations of the vector $\mathbf{r}$, respectively. Upon the formation of scalar products of equations (1.1.5.1) and (1.1.5.2) with the vectors of the second basis, and employing again the summation convention, we obtain

$$u_k(1)[\mathbf{c}_k(1)\cdot\mathbf{c}_l(2)] = u_k(2)[\mathbf{c}_k(2)\cdot\mathbf{c}_l(2)], \quad l = 1, 2, 3 \quad (1.1.5.3)$$

or

$$u_k(1)G_{kl}(12) = u_k(2)G_{kl}(22), \quad l = 1, 2, 3, \quad (1.1.5.4)$$

where $G_{kl}(12) = \mathbf{c}_k(1)\cdot\mathbf{c}_l(2)$ and $G_{kl}(22) = \mathbf{c}_k(2)\cdot\mathbf{c}_l(2)$. Similarly, if we choose the basis vectors $\mathbf{c}_l(1)$, $l = 1, 2, 3$, as the multipliers of (1.1.5.1) and (1.1.5.2), we obtain

$$u_k(1)G_{kl}(11) = u_k(2)G_{kl}(21), \quad l = 1, 2, 3, \quad (1.1.5.5)$$

where $G_{kl}(11) = \mathbf{c}_k(1)\cdot\mathbf{c}_l(1)$ and $G_{kl}(21) = \mathbf{c}_k(2)\cdot\mathbf{c}_l(1)$. Rewriting (1.1.5.4) and (1.1.5.5) in symbolic matrix notation, we have

$$\boldsymbol{u}^T(1)\boldsymbol{G}(12) = \boldsymbol{u}^T(2)\boldsymbol{G}(22), \quad (1.1.5.6)$$

leading to

$$\boldsymbol{u}^T(1) = \boldsymbol{u}^T(2)\{\boldsymbol{G}(22)[\boldsymbol{G}(12)]^{-1}\}$$

and

$$\boldsymbol{u}^T(2) = \boldsymbol{u}^T(1)\{\boldsymbol{G}(12)[\boldsymbol{G}(22)]^{-1}\}, \tag{1.1.5.7}$$

and

$$\boldsymbol{u}^T(1)\boldsymbol{G}(11) = \boldsymbol{u}^T(2)\boldsymbol{G}(21), \tag{1.1.5.8}$$

leading to

$$\boldsymbol{u}^T(1) = \boldsymbol{u}^T(2)\{\boldsymbol{G}(21)[\boldsymbol{G}(11)]^{-1}\}$$

and

$$\boldsymbol{u}^T(2) = \boldsymbol{u}^T(1)\{\boldsymbol{G}(11)[\boldsymbol{G}(21)]^{-1}\}. \tag{1.1.5.9}$$

Equations (1.1.5.7) and (1.1.5.9) are symbolic general expressions for the transformation of the coordinates of **r** from one representation to the other.

In the general case, therefore, we require the matrices of scalar products of the basis vectors, $\boldsymbol{G}(12)$ and $\boldsymbol{G}(22)$ *or* $\boldsymbol{G}(11)$ and $\boldsymbol{G}(21)$ – depending on whether the basis $\mathbf{c}_k(2)$ or $\mathbf{c}_k(1)$, $k = 1, 2, 3$, was chosen to multiply scalarly equations (1.1.5.1) and (1.1.5.2). Note, however, the following simplifications.

(i) If the bases $\mathbf{c}_k(1)$ and $\mathbf{c}_k(2)$ are mutually reciprocal, each of the matrices of mixed scalar products, $\boldsymbol{G}(12)$ and $\boldsymbol{G}(21)$, reduces to a unit matrix. In this important special case, the transformation is effected by the matrices of the metric tensors of the bases in question. This can be readily seen from equations (1.1.5.7) and (1.1.5.9), which then reduce to the relationships between the covariant and contravariant components of the same vector [see equations (1.1.4.11) and (1.1.4.12) above].

(ii) If one of the bases, say $\mathbf{c}_k(2)$, is Cartesian, its metric tensor is by definition a unit tensor, and the transformations in (1.1.5.7) reduce to

$$\boldsymbol{u}^T(1) = \boldsymbol{u}^T(2)[\boldsymbol{G}(12)]^{-1}$$

and

$$\boldsymbol{u}^T(2) = \boldsymbol{u}^T(1)\boldsymbol{G}(12). \tag{1.1.5.10}$$

The transformation matrix is now the mixed matrix of the scalar products, whether or not the basis $\mathbf{c}_k(1)$, $k = 1, 2, 3$, is also Cartesian. If, however, both bases are Cartesian, the transformation can also be interpreted as a rigid rotation of the coordinate axes (see Chapter 3.3).

It should be noted that the above transformations do not involve any shift of the origin. Transformations involving such shifts, notably the symmetry transformations of the space group, are treated rather extensively in Volume A of *International Tables for Crystallography* (2005) [see *e.g.* Part 5 there (Arnold, 2005)].

1.1.5.2. Example

This example deals with the construction of a Cartesian system in a crystal with given basis vectors of its direct lattice. We shall also require that the Cartesian system bear a clear relationship to at least one direction in each of the direct and reciprocal lattices of the crystal; this may be useful in interpreting a physical property which has been measured along a given lattice vector or which is associated with a given lattice plane. For a better consistency of notation, the Cartesian components will be denoted as contravariant.

The appropriate version of equations (1.1.5.1) and (1.1.5.2) is now

$$\mathbf{r} = x^i\mathbf{a}_i \tag{1.1.5.11}$$

and

$$\mathbf{r} = X^k\mathbf{e}_k, \tag{1.1.5.12}$$

where the Cartesian basis vectors are: $\mathbf{e}_1 = \mathbf{r}_\mathrm{L}/|\mathbf{r}_\mathrm{L}|$, $\mathbf{e}_2 = \mathbf{r}^*/|\mathbf{r}^*|$ and $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$, and the vectors $\mathbf{r}_\mathrm{L}$ and $\mathbf{r}^*$ are given by

$$\mathbf{r}_\mathrm{L} = u^i\mathbf{a}_i \text{ and } \mathbf{r}^* = h_k\mathbf{a}^k,$$

where u^i and h_k, $i, k = 1, 2, 3$, are arbitrary integers. The vectors $\mathbf{r}_\mathrm{L}$ and $\mathbf{r}^*$ must of course be chosen to be mutually perpendicular, $\mathbf{r}_\mathrm{L} \cdot \mathbf{r}^* = u^ih_i = 0$. The $X^1(X)$ axis of the Cartesian system thus coincides with a direct-lattice vector, and the $X^2(Y)$ axis is parallel to a vector in the reciprocal lattice.

Since the basis in (1.1.5.12) is a Cartesian one, the required transformations are given by equations (1.1.5.10) as

$$x^i = X^k(T^{-1})^i_k \text{ and } X^i = x^kT^i_k, \tag{1.1.5.13}$$

where $T^i_k = \mathbf{a}_k \cdot \mathbf{e}_i$, k, $i = 1, 2, 3$, form the matrix of the scalar products. If we make use of the relationships between covariant and contravariant basis vectors, and the tensor formulation of the vector product, given in Section 1.1.4 above (see also Chapter 3.1), we obtain

$$\begin{aligned} T^1_k &= \frac{1}{|\mathbf{r}_\mathrm{L}|}g_{ki}u^i \\ T^2_k &= \frac{1}{|\mathbf{r}^*|}h_k \\ T^3_k &= \frac{V}{|\mathbf{r}_\mathrm{L}||\mathbf{r}^*|}e_{kip}u^ig^{pl}h_l. \end{aligned} \tag{1.1.5.14}$$

Note that the other convenient choice, $\mathbf{e}_1 \propto \mathbf{r}^*$ and $\mathbf{e}_2 \propto \mathbf{r}_\mathrm{L}$, interchanges the first two columns of the matrix $\boldsymbol{T}$ in (1.1.5.14) and leads to a change of the signs of the elements in the third column. This can be done by writing e_{kpi} instead of e_{kip}, while leaving the rest of T^3_k unchanged.

1.1.6. Some analytical aspects of the reciprocal space

1.1.6.1. Continuous Fourier transform

Of great interest in crystallographic analyses are Fourier transforms and these are closely associated with the dual bases examined in this chapter. Thus, *e.g.*, the inverse Fourier transform of the electron-density function of the crystal

$$F(\mathbf{h}) = \int_\text{cell} \rho(\mathbf{r}) \exp(2\pi i\mathbf{h} \cdot \mathbf{r})\, \mathrm{d}^3\mathbf{r}, \tag{1.1.6.1}$$

where $\rho(\mathbf{r})$ is the electron-density function at the point **r** and the integration extends over the volume of a unit cell, is the fundamental model of the contribution of the distribution of crystalline matter to the intensity of the scattered radiation. For the conventional Bragg scattering, the function given by (1.1.6.1), and known as the *structure factor*, may assume nonzero values *only* if **h** can be represented as a reciprocal-lattice vector. Chapter 1.2 is devoted to a discussion of the structure factor of the Bragg reflection, while Chapters 4.1, 4.2 and 4.3 discuss circumstances under which the scattering need not be confined to the points of the reciprocal lattice only, and may be represented by reciprocal-space vectors with non-integral components.

1.1.6.2. Discrete Fourier transform

The electron density $\rho(\mathbf{r})$ in (1.1.6.1) is one of the most common examples of a function which has the periodicity of the crystal. Thus, for an ideal (infinite) crystal the electron density $\rho(\mathbf{r})$ can be written as

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + u\mathbf{a} + v\mathbf{b} + w\mathbf{c}), \tag{1.1.6.2}$$

and, as such, it can be represented by a three-dimensional Fourier series of the form

$$\rho(\mathbf{r}) = \sum_{\mathbf{g}} C(\mathbf{g}) \exp(-2\pi i \mathbf{g} \cdot \mathbf{r}), \qquad (1.1.6.3)$$

where the periodicity requirement (1.1.6.2) enables one to represent all the **g** vectors in (1.1.6.3) as vectors in the reciprocal lattice (see also Section 1.1.2 above). If we insert the series (1.1.6.3) in the integrand of (1.1.6.1), interchange the order of summation and integration and make use of the fact that an integral of a periodic function taken over the entire period must vanish unless the integrand is a constant, equation (1.1.6.3) reduces to the conventional form

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}), \qquad (1.1.6.4)$$

where V is the volume of the unit cell in the direct lattice and the summation ranges over all the reciprocal lattice.

Fourier transforms, discrete as well as continuous, are among the most important mathematical tools of crystallography. The discussion of their mathematical principles, the modern algorithms for their computation and their numerous applications in crystallography form the subject matter of Chapter 1.3. Many more examples of applications of Fourier methods in crystallography are scattered throughout this volume and the crystallographic literature in general.

1.1.6.3. Bloch's theorem

It is in order to mention briefly the important role of reciprocal space and the reciprocal lattice in the field of the theory of solids. At the basis of these applications is the periodicity of the crystal structure and the effect it has on the dynamics (*cf.* Chapter 4.1) and electronic structure of the crystal. One of the earliest, and still most important, theorems of solid-state physics is due to Bloch (1928) and deals with the representation of the wavefunction of an electron which moves in a periodic potential. Bloch's theorem states that:

> *The eigenstates ψ of the one-electron Hamiltonian $\mathcal{H} = (-\hbar^2/2m)\nabla^2 + U(\mathbf{r})$, where $U(\mathbf{r})$ is the crystal potential and $\mathbf{U}(\mathbf{r} + \mathbf{r}_\mathrm{L}) = U(\mathbf{r})$ for all $\mathbf{r}_\mathrm{L}$ in the Bravais lattice, can be chosen to have the form of a plane wave times a function with the periodicity of the Bravais lattice.*

Thus

$$\psi(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u(\mathbf{r}), \qquad (1.1.6.5)$$

where

$$u(\mathbf{r} + \mathbf{r}_\mathrm{L}) = u(\mathbf{r}) \qquad (1.1.6.6)$$

and **k** is the wavevector. The proof of Bloch's theorem can be found in most modern texts on solid-state physics (*e.g.* Ashcroft & Mermin, 1975). If we combine (1.1.6.5) with (1.1.6.6), an alternative form of the Bloch theorem results:

$$\psi(\mathbf{r} + \mathbf{r}_\mathrm{L}) = \exp(i\mathbf{k} \cdot \mathbf{r}_\mathrm{L}) \psi(\mathbf{r}). \qquad (1.1.6.7)$$

In the important case where the wavefunction ψ is itself periodic, *i.e.*

$$\psi(\mathbf{r} + \mathbf{r}_\mathrm{L}) = \psi(\mathbf{r}),$$

we must have $\exp(i\mathbf{k} \cdot \mathbf{r}_\mathrm{L}) = 1$. Of course, this can be so only if the wavevector **k** equals 2π times a vector in the reciprocal lattice. It is also seen from equation (1.1.6.7) that the wavevector appearing in the phase factor can be reduced to a unit cell in the reciprocal lattice (the basis vectors of which contain the 2π factor), or to the equivalent polyhedron known as the *Brillouin zone* (*e.g.* Ziman, 1969). This periodicity in reciprocal space is of prime importance in the theory of solids. Some Brillouin zones are discussed in detail in Chapter 1.5.

I wish to thank Professor D. W. J. Cruickshank for bringing to my attention the contribution of M. von Laue (Laue, 1914), who was the first to introduce general reciprocal bases to crystallography.

References

Arnold, H. (2005). *Transformations in crystallography*. In *International Tables for Crystallography*, Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, Part 5. Heidelberg: Springer.

Ashcroft, N. W. & Mermin, N. D. (1975). *Solid State Physics.* Philadelphia: Saunders College.

Bloch, F. (1928). *Über die Quantenmechanik der Elektronen in Kristallgittern. Z. Phys.* **52**, 555–600.

Buerger, M. J. (1941). *X-ray Crystallography.* New York: John Wiley.

Buerger, M. J. (1959). *Crystal Structure Analysis.* New York: John Wiley.

Ewald, P. P. (1913). *Zur Theorie der Interferenzen der Röntgenstrahlen in Kristallen. Phys. Z.* **14**, 465–472.

Ewald, P. P. (1921). *Das reziproke Gitter in der Strukturtheorie. Z. Kristallogr.* **56**, 129–156.

International Tables for Crystallography (2005). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn. Heidelberg: Springer.

International Tables for Crystallography (2004). Vol. C, *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

Koch, E. (2004). In *International Tables for Crystallography*, Vol. C, *Mathematical, Physical and Chemical Tables*, edited by E. Prince, Chapters 1.1 and 1.2. Dordrecht: Kluwer Academic Publishers.

Laue, M. (1914). *Die Interferenzerscheinungen an Röntgenstrahlen, hervorgerufen durch das Raumgitter der Kristalle. Jahrb. Radioakt. Elektron.* **11**, 308–345.

Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures.* London: Bell.

Nespolo, M. (2015). *The ash heap of crystallography: restoring forgotten basic knowledge. J. Appl. Cryst.* **48**, 1290–1298.

Patterson, A. L. (1967). In *International Tables for X-ray Crystallography*, Vol. II, *Mathematical Tables*, edited by J. S. Kasper & K. Lonsdale, pp. 5–83. Birmingham: Kynoch Press.

Sands, D. E. (1982). *Vectors and Tensors in Crystallography.* New York: Addison-Wesley.

Schomaker, V. & Trueblood, K. N. (1968). *On the rigid-body motion of molecules in crystals. Acta Cryst.* B**24**, 63–76.

Shmueli, U. (2007). *Theories and Techniques of Crystal Structure Determination*, Section 1.2. Oxford University Press.

Wilson, E. B. (1901). *Vector Analysis.* New Haven: Yale University Press.

Ziman, J. M. (1969). *Principles of the Theory of Solids.* Cambridge University Press.

International Tables for Crystallography (2010). Vol. B, Chapter 1.2, pp. 10–23.

1.2. The structure factor

By P. Coppens

1.2.1. Introduction

The *structure factor* is the central concept in structure analysis by diffraction methods. Its modulus is called the *structure amplitude*. The structure amplitude is a function of the indices of the set of scattering planes h, k and l, and is defined as the amplitude of scattering by the contents of the crystallographic unit cell, expressed in units of scattering. For X-ray scattering, that unit is the scattering by a single electron (2.82×10^{-15} m), while for neutron scattering by atomic nuclei, the unit of scattering length of 10^{-14} m is commonly used. The complex form of the structure factor means that the phase of the scattered wave is not simply related to that of the incident wave. However, the observable, which is the scattered intensity, must be real. It is proportional to the square of the scattering amplitude (see, *e.g.*, Lipson & Cochran, 1966).

The structure factor is directly related to the distribution of scattering matter in the unit cell which, in the X-ray case, is the electron distribution, time-averaged over the vibrational modes of the solid.

In this chapter we will discuss structure-factor expressions for X-ray and neutron scattering, and, in particular, the modelling that is required to obtain an analytical description in terms of the features of the electron distribution and the vibrational displacement parameters of individual atoms. We concentrate on the most basic developments; for further details the reader is referred to the cited literature.

1.2.2. General scattering expression for X-rays

The total scattering of X-rays contains both elastic and inelastic components. Within the first-order Born approximation (Born, 1926) it has been treated by several authors (*e.g.* Waller & Hartree, 1929; Feil, 1977) and is given by the expression

$$I_{\rm total}(\mathbf{S}) = I_{\rm classical} \sum_n \left| \int \psi_n^* \exp(2\pi i \mathbf{S}\cdot\mathbf{r}_j)\psi_0 \, {\rm d}\mathbf{r} \right|^2, \qquad (1.2.2.1)$$

where $I_{\rm classical}$ is the classical Thomson scattering of an X-ray beam by a free electron, which is equal to $(e^2/mc^2)^2(1+\cos^2 2\theta)/2$ for an unpolarized beam of unit intensity, ψ is the n-electron space-wavefunction expressed in the $3n$ coordinates of the electrons located at $\mathbf{r}_j$ and the integration is over the coordinates of all electrons. $\mathbf{S}$ is the scattering vector of length $2\sin\theta/\lambda$.

The coherent elastic component of the scattering, in units of the scattering of a free electron, is given by

$$I_{\rm coherent,\ elastic}(\mathbf{S}) = \left| \int \psi_0^* \right| \sum_j \exp(2\pi i \mathbf{S}\cdot\mathbf{r}_j) |\psi_0 \, {\rm d}\mathbf{r}|^2. \qquad (1.2.2.2)$$

If integration is performed over all coordinates but those of the jth electron, one obtains after summation over all electrons

$$I_{\rm coherent,\ elastic}(\mathbf{S}) = |\int \rho(\mathbf{r}) \exp(2\pi i \mathbf{S}\cdot\mathbf{r}) \, {\rm d}\mathbf{r}|^2, \qquad (1.2.2.3)$$

where $\rho(\mathbf{r})$ is the electron distribution. The scattering amplitude $A(\mathbf{S})$ is then given by

$$A(\mathbf{S}) = \int \rho(\mathbf{r}) \exp(2\pi i \mathbf{S}\cdot\mathbf{r}) \, {\rm d}\mathbf{r} \qquad (1.2.2.4a)$$

or

$$A(\mathbf{S}) = \hat{F}\{\rho(\mathbf{r})\}, \qquad (1.2.2.4b)$$

where $\hat{F}$ is the Fourier transform operator.

1.2.3. Scattering by a crystal: definition of a structure factor

In a crystal of infinite size, $\rho(\mathbf{r})$ is a three-dimensional periodic function, as expressed by the convolution

$$\rho_{\rm crystal}(\mathbf{r}) = \sum_n \sum_m \sum_p \rho_{\rm unit\ cell}(\mathbf{r}) * \delta(\mathbf{r} - n\mathbf{a} - m\mathbf{b} - p\mathbf{c}), \qquad (1.2.3.1)$$

where n, m and p are integers, and δ is the Dirac delta function.

Thus, according to the Fourier convolution theorem,

$$A(\mathbf{S}) = \hat{F}\{\rho(\mathbf{r})\} = \sum_n \sum_m \sum_p \hat{F}\{\rho_{\rm unit\ cell}(\mathbf{r})\} \hat{F}\{\delta(\mathbf{r} - n\mathbf{a} - m\mathbf{b} - p\mathbf{c})\}, \qquad (1.2.3.2)$$

which gives

$$A(\mathbf{S}) = \hat{F}\{\rho_{\rm unit\ cell}(\mathbf{r})\} \sum_h \sum_k \sum_l \delta(\mathbf{S} - h\mathbf{a}^* - k\mathbf{b}^* - l\mathbf{c}^*). \qquad (1.2.3.3)$$

Expression (1.2.3.3) is valid for a crystal with a very large number of unit cells, in which particle-size broadening is negligible. Furthermore, it does not account for multiple scattering of the beam within the crystal. Because of the appearance of the delta function, (1.2.3.3) implies that $\mathbf{S} = \mathbf{H}$ with $\mathbf{H} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$.

The first factor in (1.2.3.3), the scattering amplitude of one unit cell, is defined as the structure factor F:

$$F(\mathbf{H}) = \hat{F}\{\rho_{\rm unit\ cell}(\mathbf{r})\} = \int_{\rm unit\ cell} \rho(\mathbf{r}) \exp(2\pi i \mathbf{H}\cdot\mathbf{r}) \, {\rm d}\mathbf{r}. \qquad (1.2.3.4)$$

1.2.4. The isolated-atom approximation in X-ray diffraction

To a reasonable approximation, the unit-cell density can be described as a superposition of isolated, spherical atoms located at $\mathbf{r}_j$.

$$\rho_{\rm unit\ cell}(\mathbf{r}) = \sum_j \rho_{{\rm atom},j}(\mathbf{r}) * \delta(\mathbf{r} - \mathbf{r}_j). \qquad (1.2.4.1)$$

Substitution in (1.2.3.4) gives

$$F(\mathbf{H}) = \sum_j \hat{F}\{\rho_{{\rm atom},j}\} \hat{F}\{\delta(\mathbf{r} - \mathbf{r}_j)\} = \sum_j f_j \exp(2\pi i \mathbf{H}\cdot\mathbf{r}_j) \qquad (1.2.4.2a)$$

or

$$\begin{aligned} F(h,k,l) &= \sum_j f_j \exp 2\pi i(hx_j + ky_j + lz_j) \\ &= \sum_j f_j \{\cos 2\pi(hx_j + ky_j + lz_j) \\ &\quad + i \sin 2\pi(hx_j + ky_j + lz_j)\}. \end{aligned} \qquad (1.2.4.2b)$$

$f_j(S)$, the spherical atomic scattering factor, or form factor, is the Fourier transform of the spherically averaged atomic density $\rho_j(r)$, in which the polar coordinate r is relative to the nuclear position. $f_j(S)$ can be written as (James, 1982)

Copyright © 2010 International Union of Crystallography

$$f_j(S) = \int_{\text{atom}} \rho_j(r) \exp(2\pi i \mathbf{S}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r}$$
$$= \int_{\upsilon=0}^{\pi}\int_{\varphi=0}^{2\pi}\int_{r=0}^{\infty} \rho_j(r)\exp(2\pi i S r\cos\vartheta) r^2 \sin\vartheta\,\mathrm{d}r\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$$
$$= \int_0^\infty 4\pi r^2 \rho_j(r)\frac{\sin 2\pi S r}{2\pi S r}\,\mathrm{d}r \equiv \int_0^\infty 4\pi r^2\rho_j(r) j_0(2\pi S r)\,\mathrm{d}r$$
$$\equiv \langle j_0\rangle, \qquad (1.2.4.3)$$

where $j_0(2\pi Sr)$ is the zero-order spherical Bessel function.

$\rho_j(r)$ represents either the static or the dynamic density of atom j. In the former case, the effect of thermal motion, treated in Section 1.2.9 and following, is not included in the expression.

When scattering is treated in the second-order Born approximation, additional terms occur which are in particular of importance for X-ray wavelengths with energies close to absorption edges of atoms, where the participation of free and bound excited states in the scattering process becomes very important, leading to *resonance scattering*. [Resonance scattering is referred to as anomalous scattering in the older literature, but this misnomer is avoided in the current chapter.] Inclusion of such contributions leads to two extra terms, which are both wavelength- and scattering-angle-dependent:

$$f_j(S,\lambda) = f_j^0(S) + f_j'(S,\lambda) + i f_j''(S,\lambda). \qquad (1.2.4.4)$$

The treatment of resonance effects is beyond the scope of this chapter. We note however (*a*) that to a reasonable approximation the S-dependence of f_j' and f_j'' can be neglected, (*b*) that f_j' and f_j'' are not independent, but related through the Kramers–Kronig transformation, and (*c*) that in an anisotropic environment the atomic scattering factor becomes anisotropic, and accordingly is described as a tensor property. Detailed descriptions and appropriate references can be found in Materlick *et al.* (1994) and in Section 4.2.6 of *IT* C (2004).

The structure-factor expressions (1.2.4.2) can be simplified when the crystal class contains nontrivial symmetry elements. For example, when the origin of the unit cell coincides with a centre of symmetry ($x, y, z \rightarrow -x, -y, -z$) the sine term in (1.2.4.2*b*) cancels when the contributions from the symmetry-related atoms are added, leading to the expression

$$F(\mathbf{H}) = 2\sum_{j=1}^{N/2} f_j \cos 2\pi(hx_j + ky_j + lz_j), \qquad (1.2.4.2c)$$

where the summation is over the unique half of the unit cell only.

Further simplifications occur when other symmetry elements are present. They are treated in Chapter 1.4, which also contains a complete list of symmetry-specific structure-factor expresions valid in the spherical-atom isotropic-temperature-factor approximation.

1.2.5. Scattering of thermal neutrons

1.2.5.1. Nuclear scattering

The scattering of neutrons by atomic nuclei is described by the atomic scattering length b, related to the total cross section σ_{total} by the expression $\sigma_{\text{total}} = 4\pi b^2$. At present, there is no theory of nuclear forces which allows calculation of the scattering length, so that experimental values are to be used. Two types of nuclei can be distinguished (Squires, 1978). In the first type, the scattering is a resonance phenomenon and is associated with the formation of a compound nucleus (consisting of the original nucleus plus a neutron) with an energy close to that of an excited state. In the second type, the compound nucleus is not near an excited state and the scattering length is essentially real and independent of the energy of the incoming neutron. In either case, b is independent of the Bragg angle θ, unlike the X-ray form factor, since the nuclear dimensions are very small relative to the wavelength of thermal neutrons.

The scattering length is not the same for different isotopes of an element. A random distribution of isotopes over the sites occupied by that element leads to an incoherent contribution, such that effectively $\sigma_{\text{total}} = \sigma_{\text{coherent}} + \sigma_{\text{incoherent}}$. Similarly for nuclei with nonzero spin, a spin incoherent scattering occurs as the spin states are, in general, randomly distributed over the sites of the nuclei.

For free or loosely bound nuclei, the scattering length is modified by $b_{\text{free}} = [M/(m+M)]b$, where M is the mass of the nucleus and m is the mass of the neutron. This effect is of consequence only for the lightest elements. It can, in particular, be of significance for hydrogen atoms. With this in mind, the structure-factor expression for elastic scattering can be written as

$$F(\mathbf{H}) = \sum_j b_{j,\,\text{coherent}} \exp 2\pi i(hx_j + ky_j + lz_j) \qquad (1.2.4.2d)$$

by analogy to (1.2.4.2*b*).

1.2.5.2. Magnetic scattering

The interaction between the magnetic moments of the neutron and the unpaired electrons in solids leads to magnetic scattering. The total elastic scattering including both the nuclear and magnetic contributions is given by

$$|F(\mathbf{H})|^2_{\text{total}} = |F_N(\mathbf{H}) + \mathbf{Q}(\mathbf{H})\cdot\hat{\boldsymbol{\lambda}}|^2, \qquad (1.2.5.1a)$$

where the unit vector $\hat{\boldsymbol{\lambda}}$ describes the polarization vector for the neutron spin, $F_N(\mathbf{H})$ is given by (1.2.4.2*b*) and $\mathbf{Q}$ is defined by

$$\mathbf{Q} = \frac{mc}{eh}\int \widehat{\mathbf{H}} \times [\mathbf{M}(\mathbf{r}) \times \widehat{\mathbf{H}}] \exp(2\pi i \mathbf{H}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r}. \qquad (1.2.5.2a)$$

$\mathbf{M}(\mathbf{r})$ is the vector field describing the electron-magnetization distribution and $\widehat{\mathbf{H}}$ is a unit vector parallel to $\mathbf{H}$.

$\mathbf{Q}$ is thus proportional to the projection of $\mathbf{M}$ onto a direction orthogonal to $\mathbf{H}$ in the plane containing $\mathbf{M}$ and $\mathbf{H}$. The magnitude of this projection depends on $\sin\alpha$, where α is the angle between $\mathbf{Q}$ and $\mathbf{H}$, which prevents magnetic scattering from being a truly three-dimensional probe. If all moments $\mathbf{M}(\mathbf{r})$ are collinear, as may be achieved in paramagnetic materials by applying an external field, and for the maximum signal ($\mathbf{H}$ orthogonal to $\mathbf{M}$), (1.2.5.2*a*) becomes

$$\mathbf{Q} = \mathbf{M}(\mathbf{H}) = \frac{mc}{eh}\int \mathbf{M}(\mathbf{r}) \exp(2\pi i \mathbf{H}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} \qquad (1.2.5.2b)$$

and (1.2.5.1*a*) gives

$$|F|^2_{\text{total}} = |F_N(\mathbf{H}) - M(\mathbf{H})|^2 \qquad (1.2.5.1b)$$

and

$$|F|^2_{\text{total}} = |F_N(\mathbf{H}) + M(\mathbf{H})|^2$$

for neutrons parallel and antiparallel to $\mathbf{M}(\mathbf{H})$, respectively.

1.2.6. Effect of bonding on the atomic electron density within the spherical-atom approximation: the kappa formalism

A first improvement beyond the isolated-atom formalism is to allow for changes in the radial dependence of the atomic electron distribution.

Such changes may be due to electronegativity differences which lead to the transfer of electrons between the valence shells of different atoms. The electron transfer introduces a change in

the screening of the nuclear charge by the electrons and therefore affects the radial dependence of the atomic electron distribution (Coulson, 1961). A change in radial dependence of the density may also occur in a purely covalent bond, as, for example, in the H_2 molecule (Ruedenberg, 1962). It can be expressed as

$$\rho'_{\rm valence}(r) = \kappa^3 \rho_{\rm valence}(\kappa r) \tag{1.2.6.1}$$

(Coppens *et al.*, 1979), where ρ' is the modified density and κ is an expansion/contraction parameter, which is > 1 for valence-shell contraction and < 1 for expansion. The κ^3 factor results from the normalization requirement.

The valence density is usually defined as the outer electron shell from which charge transfer occurs. The inner or core electrons are much less affected by the change in occupancy of the outer shell and, in a reasonable approximation, retain their radial dependence.

The corresponding structure-factor expression is

$$F(\mathbf{H}) = \sum_j [\{P_{j,\,\rm core} f_{j,\,\rm core}(H) + P_{j,\,\rm valence} f_{j,\,\rm valence}(H/\kappa)\} \times \exp(2\pi i \mathbf{H}\cdot\mathbf{r}_j)], \tag{1.2.6.2}$$

where $P_{j,\,\rm core}$ and $P_{j,\,\rm valence}$ are the number of electrons (not necessarily integral) in the core and valence shell, respectively, and the atomic scattering factors $f_{j,\,\rm core}$ and $f_{j,\,\rm valence}$ are normalized to one electron. Here and in the following sections, the resonant-scattering contributions are incorporated in the core scattering.

1.2.7. Beyond the spherical-atom description: the atom-centred spherical harmonic expansion

1.2.7.1. Direct-space description of aspherical atoms

Even though the spherical-atom approximation is often adequate, atoms in a crystal are in a nonspherical environment; therefore, an accurate description of the atomic electron density requires nonspherical density functions. In general, such density functions can be written in terms of the three polar coordinates r, θ and φ. Under the assumption that the radial and angular parts can be separated, one obtains for the density function:

$$\Phi(r,\theta,\varphi) = R(r)\Theta(\theta,\varphi). \tag{1.2.7.1}$$

The angular functions Θ are based on the spherical harmonic functions Y_{lm} defined by

$$Y_{lm}(\theta,\varphi) = (-1)^m \left[\left(\frac{2l+1}{4\pi}\right)\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} P_l^m(\cos\theta)\exp(im\varphi), \tag{1.2.7.2a}$$

with $-l \le m \le l$, where $P_l^m(\cos\theta)$ are the associated Legendre polynomials (see Arfken, 1970).

$$P_l^m(x) = (1-x^2)^{|m|/2}\frac{{\rm d}^{|m|}P_l(x)}{{\rm d}x^{|m|}},$$

$$P_l(x) = \frac{1}{l!2^l}\frac{{\rm d}^l}{{\rm d}x^l}\left[(x^2-1)^l\right].$$

The real spherical harmonic functions y_{lmp}, $0 \le m \le l$, $p = +$ or $-$ are obtained as a linear combination of Y_{lm}:

$$y_{lm+}(\theta,\psi) = \left[\frac{(2l+1)(l-|m|)!}{2\pi(1+\delta_{m0})(l+|m|)!}\right]^{1/2} P_l^m(\cos\theta)\cos m\varphi = N_{lm}P_l^m(\cos\theta)\cos m\varphi = (-1)^m(Y_{lm}+Y_{l,\,-m}) \tag{1.2.7.2b}$$

and

$$y_{lm-}(\theta,\psi) = N_{lm}P_l^m(\cos\theta)\sin m\varphi = (-1)^m(Y_{lm}-Y_{l,\,-m})/(2i). \tag{1.2.7.2c}$$

The normalization constants N_{lm} are defined by the conditions

$$\int y_{lmp}^2\ {\rm d}\Omega = 1, \tag{1.2.7.3a}$$

which are appropriate for normalization of wavefunctions. An alternative definition is used for charge-density basis functions:

$$\int |d_{lmp}|\ {\rm d}\Omega = 2 \text{ for } l>0 \text{ and } \int |d_{lmp}|\ {\rm d}\Omega = 1 \text{ for } l = 0. \tag{1.2.7.3b}$$

The functions y_{lmp} and d_{lmp} differ only in the normalization constants. For the spherically symmetric function d_{00}, a population parameter equal to one corresponds to the function being populated by one electron. For the nonspherical functions with $l > 0$, a population parameter equal to one implies that one electron has shifted from the negative to the positive lobes of the function.

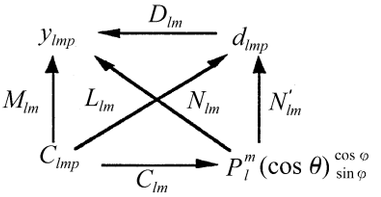
The functions y_{lmp} and d_{lmp} can be expressed in Cartesian coordinates, such that

$$y_{lmp} = M_{lm}c_{lmp} \tag{1.2.7.4a}$$

and

$$d_{lmp} = L_{lm}c_{lmp}, \tag{1.2.7.4b}$$

where the c_{lmp} are Cartesian functions. The relations between the various definitions of the real spherical harmonic functions are summarized by



in which the direction of the arrows and the corresponding conversion factors X_{lm} define expressions of the type (1.2.7.4). The expressions for c_{lmp} with $l \le 4$ are listed in Table 1.2.7.1, together with the normalization factors M_{lm} and L_{lm}.

The spherical harmonic functions are mutually orthogonal and form a complete set, which, if taken to sufficiently high order, can be used to describe any arbitrary angular function.

The spherical harmonic functions are often referred to as multipoles since each represents the components of the charge distribution $\rho(\mathbf{r})$, which gives nonzero contribution to the integral $\Theta_{lmp} = \int \rho(\mathbf{r}) c_{lmp} r^l\ {\rm d}\mathbf{r}$, where Θ_{lmp} is an electrostatic multipole moment. Terms with increasing l are referred to as *monopolar* ($l = 0$), *dipolar* ($l = 1$), *quadrupolar* ($l = 2$), *octapolar* ($l = 3$), *hexadecapolar* ($l = 4$), *triacontadipolar* ($l = 5$) and *hexacontatetrapolar* ($l = 6$).

Site-symmetry restrictions for the real spherical harmonics as given by Kara & Kurki-Suonio (1981) are summarized in Table 1.2.7.2.

In cubic space groups, the spherical harmonic functions as defined by equations (1.2.7.2) are no longer linearly independent. The appropriate basis set for this symmetry consists of the '*Kubic harmonics*' of Von der Lage & Bethe (1947). Some low-order terms are listed in Table 1.2.7.3. Both wavefunction and density-function normalization factors are specified in Table 1.2.7.3.

A related basis set of angular functions has been proposed by Hirshfeld (1977). They are of the form $\cos^n\theta_k$, where θ_k is the angle with a specified set of $(n+1)(n+2)/2$ polar axes. The Hirshfeld functions are identical to a sum of spherical harmonics

Table 1.2.7.1. *Real spherical harmonic functions (x, y, z are direction cosines)*

l	Symbol	C†	Angular function, c_{lmp}‡	Normalization for wavefunctions, M_{lmp}§		Normalization for density functions, L_{lmp}¶	
				Expression	Numerical value	Expression	Numerical value
0	00	1	1	$(1/4\pi)^{1/2}$	0.28209	$1/4\pi$	0.07958
1	11+	1	x	$(3/4\pi)^{1/2}$	0.48860	$1/\pi$	0.31831
	11−	1	y				
	10	1	z				
2	20	1/2	$3z^2-1$	$(5/16\pi)^{1/2}$	0.31539	$\frac{3\sqrt{3}}{8\pi}$	0.20675
	21+	3	xz	$(15/4\pi)^{1/2}$	1.09255	3/4	0.75
	21−	3	yz				
	22+	6	$(x^2-y^2)/2$				
	22−	6	xy				
3	30	1/2	$5z^3-3z$	$(7/16\pi)^{1/2}$	0.37318	$\frac{10}{13\pi}$	0.24485
	31+	3/2	$x[5z^2-1]$	$(21/32\pi)^{1/2}$	0.45705	$[\mathrm{ar}+(14/5)-(\pi/4)]^{-1}$††	0.32033
	31−	3/2	$y[5z^2-1]$				
	32+	15	$(x^2-y^2)z$	$(105/16\pi)^{1/2}$	1.44531	1	1
	32−	15	$2xyz$				
	33+	15	x^3-3xy^2	$(35/32\pi)^{1/2}$	0.59004	$4/3\pi$	0.42441
	33−	15	$-y^3+3x^2y$				
4	40	1/8	$35z^4-30z^2+3$	$(9/256\pi)^{1/2}$	0.10579	‡‡	0.06942
	41+	5/2	$x[7z^3-3z]$	$(45/32\pi)^{1/2}$	0.66905	$\frac{735}{512\sqrt{7}+196}$	0.47400
	41−	5/2	$y[7z^3-3z]$				
	42+	15/2	$(x^2-y^2)[7z^2-1]$	$(45/64\pi)^{1/2}$	0.47309	$\frac{105\sqrt{7}}{4(136+28\sqrt{7})}$	0.33059
	42−	15/2	$2xy[7z^2-1]$				
	43+	105	$(x^3-3xy^2)z$	$(315/32\pi)^{1/2}$	1.77013	5/4	1.25
	43−	105	$(-y^3+3x^2y)z$				
	44+	105	$x^4-6x^2y^2+y^4$	$(315/256\pi)^{1/2}$	0.62584	15/32	0.46875
	44−	105	$4x^3y-4xy^3$				
5	50	1/8	$63z^5-70z^3-15z$	$(11/256\pi)^{1/2}$	0.11695	—	0.07674
	51+	15/8	$(21z^4-14z^2+1)x$	$(165/256\pi)^{1/2}$	0.45295	—	0.32298
	51−		$(21z^4-14z^2+1)y$				
	52+	105/2	$(3z^3-z)(x^2-y^2)$	$(1155/64\pi)^{1/2}$	2.39677	—	1.68750
	52−		$2xy(3z^3-z)$				
	53+	105/2	$(9z^2-1)(x^3-3xy^2)$	$(385/512\pi)^{1/2}$	0.48924	—	0.34515
	53−		$(9z^2-1)(3x^2y-y^3)$				
	54+	945	$z(x^4-6x^2y^2+y^4)$	$(3465/256\pi)^{1/2}$	2.07566	—	1.50000
	54−		$z(4x^3y-4xy^3)$				
	55+	945	$x^5-10x^3y^2+5xy^4$	$(693/512\pi)^{1/2}$	0.65638	—	0.50930
	55−		$5x^4y-10x^2y^3+y^5$				

with $l = n, n-2, n-4, \ldots (0, 1)$ for $n > 1$, as shown elsewhere (Hirshfeld, 1977).

The radial functions $R(r)$ can be selected in different manners. Several choices may be made, such as

$$R_l(r) = \frac{\zeta^{n_l+3}}{(n_l+2)!} r^{n(l)} \exp(-\zeta_l r) \qquad \text{(Slater type function)}, \tag{1.2.7.5a}$$

where the coefficient n_l may be selected by examination of products of hydrogenic orbitals which give rise to a particular multipole (Hansen & Coppens, 1978). Values for the exponential coefficient ζ_l may be taken from energy-optimized coefficients for isolated atoms available in the literature (Clementi & Raimondi, 1963). A standard set has been proposed by Hehre *et al.* (1969). In the bonded atom, such values are affected by changes in nuclear screening due to migrations of charge, as described in part by equation (1.2.6.1).

Other alternatives are:

$$R_l(r) = \frac{\alpha^{n+1}}{n!} r^n \exp(-\alpha r^2) \qquad \text{(Gaussian function)} \tag{1.2.7.5b}$$

Table 1.2.7.1 (*cont.*)

l	Symbol	C†	Angular function, c_{lmp}‡	Normalization for wavefunctions, M_{lmp}§		Normalization for density functions, L_{lmp}¶	
				Expression	Numerical value	Expression	Numerical value
6	60	1/16	$231z^6 - 315z^4 + 105z^2 - 5$	$(13/1024\pi)^{1/2}$	0.06357	—	0.04171
	61+ 61−	21/8	$(33z^5 - 30z^3 + 5z)x$ $(33z^5 - 30z^3 + 5z)y$	$(273/256\pi)^{1/2}$	0.58262	—	0.41721
	62+ 62−	105/8	$(33z^4 - 18z^2 + 1)(x^2 - y^2)$ $2xy(33z^4 - 18z^2 + 1)$	$(1365/2048\pi)^{1/2}$	0.46060	—	0.32611
	63+ 63−	315/2	$(11z^3 - 3z)(x^3 - 3xy^2)$ $(11z^3 - 3z)(3x^2y - 3y)$	$(1365/512\pi)^{1/2}$	0.92121	—	0.65132
	64+ 64−	945/2	$(11z^2 - 1)(x^4 - 6x^2y^2 + y^4)$ $(11z^2 - 1)(4x^3y - 4xy^3)$	$(819/1024\pi)^{1/2}$	0.50457	—	0.36104
	65+ 65−	10395	$z(x^5 - 10x^3y^2 + 5xy^4)$ $z(5x^4y - 10x^2y^3 + y^5)$	$(9009/512\pi)^{1/2}$	2.36662	—	1.75000
	66+ 66−	10395	$x^6 - 15x^4y^2 + 15x^2y^4 - y^6$ $6x^5y - 20x^3y^3 + 6xy^5$	$(3003/2048\pi)^{1/2}$	0.68318	—	0.54687
7	70	1/16	$429z^7 - 693z^5 + 315z^3 - 35z$	$(15/1024\pi)^{1/2}$	0.06828	—	0.04480
	71+ 71−	7/16	$(429z^6 - 495z^4 + 135z^2 - 5)x$ $(429z^6 - 495z^4 + 135z^2 - 5)y$	$(105/4096\pi)^{1/2}$	0.09033	—	0.06488
	72+ 72−	63/8	$(143z^5 - 110z^3 + 15z)(x^2 - y^2)$ $2xy(143z^5 - 110z^3 + 15z)$	$(315/2048\pi)^{1/2}$	0.22127	—	0.15732
	73+ 73−	315/8	$(143z^4 - 66z^2 + 3)(x^3 - 3xy^2)$ $(143z^4 - 66z^2 + 3)(3x^2y - y^3)$	$(315/4096\pi)^{1/2}$	0.15646	—	0.11092
	74+ 74−	3465/2	$(13z^3 - 3z)(x^4 - 6x^2y^2 + y^4)$ $(13z^3 - 3z)(4x^3y - 4xy^3)$	$(3465/1024\pi)^{1/2}$	1.03783	—	0.74044
	75+ 75−	10395/2	$(13z^3 - 1)(x^5 - 10x^3y^2 + 5xy^4)$ $(13z^3 - 1)(5x^4y - 10x^2y^3 + y^5)$	$(3465/4096\pi)^{1/2}$	0.51892	—	0.37723
	76+ 76−	135135	$z(x^6 - 15x^4y^2 + 15x^2y^4 - y^6)$ $z(6x^5y + 20x^3y^3 - 6xy^5)$	$(45045/2048\pi)^{1/2}$	2.6460	—	2.00000
	77+ 77−	135135	$x^7 - 21x^5y^2 + 35x^3y^4 - 7xy^6$ $7x^6y - 35x^4y^3 + 21x^2y^5 - y^7$	$(6435/4096\pi)^{1/2}$	0.70716	—	0.58205

† Common factor such that $C_{lm}c_{lmp} = P_l^m(\cos\theta)^{\cos m\varphi}_{\sin m\varphi}$. ‡ $x = \sin\theta\cos\varphi$, $y = \sin\theta\sin\varphi$, $z = \cos\theta$. § As defined by $y_{lmp} = M_{lmp}c_{lmp}$ where c_{lmp} are Cartesian functions. ¶ Paturle & Coppens (1988), as defined by $d_{lmp} = L_{lmp}c_{lmp}$ where c_{lmp} are Cartesian functions. †† ar = arctan (2). ‡‡ $N_{\rm ang} = \{(14A_-^5 - 14A_+^5 + 20A_+^3 - 20A_-^3 + 6A_- - 6A_+)2\pi\}^{-1}$ where $A_\pm = [(30 \pm \sqrt{480})/70]^{1/2}$.

or

$$R_l(r) = r^l L_n^{2l+2}(\gamma r)\exp\left(-\frac{\gamma r}{2}\right) \quad \text{(Laguerre function)}, \tag{1.2.7.5c}$$

where L is a Laguerre polynomial of order n and degree $(2l + 2)$.

In summary, in the multipole formalism the atomic density is described by

$$\rho_{\rm atomic}(\mathbf{r}) = P_c\rho_{\rm core} + P_v\kappa^3\rho_{\rm valence}(\kappa r) + \sum_{l=0}^{l_{\max}}\kappa'^3 R_l(\kappa' r)\sum_{m=0}^{l}\sum_p P_{lmp}d_{lmp}(\mathbf{r}/r), \tag{1.2.7.6}$$

in which the leading terms are those of the kappa formalism [expressions (1.2.6.1), (1.2.6.2)]; the subscript p is either + or −.

The expansion in (1.2.7.6) is frequently truncated at the hexadecapolar ($l = 4$) level. For atoms at positions of high site symmetry the first allowed functions may occur at higher l values. For trigonally bonded atoms in organic molecules the $l = 3$ terms are often found to be the most significantly populated deformation functions.

1.2.7.2. Reciprocal-space description of aspherical atoms

The aspherical-atom form factor is obtained by substitution of (1.2.7.6) in expression (1.2.4.3*a*):

$$f_j(\mathbf{S}) = \int \rho_j(\mathbf{r})\exp(2\pi i\mathbf{S}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r}. \tag{1.2.4.3a}$$

In order to evaluate the integral, the scattering operator $\exp(2\pi i\mathbf{S}\cdot\mathbf{r})$ must be written as an expansion of products of spherical harmonic functions. In terms of the complex spherical harmonic functions, the appropriate expression is (Weiss & Freeman, 1959; Cohen-Tannoudji *et al.*, 1977)

$$\exp(2\pi i\mathbf{S}\cdot\mathbf{r}) = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l} i^l j_l(2\pi Sr)Y_{lm}(\theta,\varphi)Y_{lm}^*(\beta,\gamma). \tag{1.2.7.7a}$$

Table 1.2.7.2. *Index-picking rules of site-symmetric spherical harmonics (Kara & Kurki-Suonio, 1981)*

λ, μ and j are integers.

Symmetry	Choice of coordinate axes	Indices of allowed y_{lmp}, d_{lmp}
1	Any	All $(l, m, \pm)$
$\bar{1}$	Any	$(2\lambda, m, \pm)$
2	$2 \parallel z$	$(l, 2\mu, \pm)$
m	$m \perp z$	$(l, l-2j, \pm)$
$2/m$	$2 \parallel z, m \perp z$	$(2\lambda, 2\mu, \pm)$
222	$2 \parallel z, 2 \parallel y$	$(2\lambda, 2\mu, +)$, $(2\lambda+1, 2\mu, -)$
$mm2$	$2 \parallel z, m \perp y$	$(l, 2\mu, +)$
mmm	$m \perp z, m \perp y, m \perp x$	$(2\lambda, 2\mu, +)$
4	$4 \parallel z$	$(l, 4\mu, \pm)$
$\bar{4}$	$\bar{4} \parallel z$	$(2\lambda, 4\mu, \pm)$, $(2\lambda+1, 4\mu+2, \pm)$
$4/m$	$4 \parallel z, m \perp z$	$(2\lambda, 4\mu, \pm)$
422	$4 \parallel z, 2 \parallel y$	$(2\lambda, 4\mu, +)$, $(2\lambda+1, 4\mu, -)$
$4mm$	$4 \parallel z, m \perp y$	$(l, 4\mu, +)$
$\bar{4}2m$	$\bar{4} \parallel z, 2 \parallel x$	$(2\lambda, 4\mu, +)$, $(2\lambda+1, 4\mu+2, -)$
	$m \perp y$	$(2\lambda, 4\mu, +)$, $(2\lambda+1, 4\mu+2, +)$
$4/mmm$	$4 \parallel z, m \perp z, m \perp x$	$(2\lambda, 4\mu, +)$
3	$3 \parallel z$	$(l, 3\mu, \pm)$
$\bar{3}$	$\bar{3} \parallel z$	$(2\lambda, 3\mu, \pm)$
32	$3 \parallel z, 2 \parallel y$	$(2\lambda, 3\mu, +)$, $(2\lambda+1, 3\mu, -)$
	$2 \parallel x$	$(3\mu+2j, 3\mu, +)$, $(3\mu+2j+1, 3\mu, -)$
$3m$	$3 \parallel z, m \perp y$	$(l, 3\mu, +)$
	$m \perp x$	$(l, 6\mu, +)$, $(l, 6\mu+3, -)$
$\bar{3}m$	$\bar{3} \parallel z, m \perp y$	$(2\lambda, 3\mu, +)$
	$m \perp x$	$(2\lambda, 6\mu, +)$, $(2\lambda, 6\mu+3, -)$
6	$6 \parallel z$	$(l, 6\mu, \pm)$
$\bar{6}$	$\bar{6} \parallel z$	$(2\lambda, 6\mu, \pm)$, $(2\lambda+1, 6\mu+3, \pm)$
$6/m$	$6 \parallel z, m \perp z$	$(2\lambda, 6\mu, \pm)$
622	$6 \parallel z, 2 \parallel y$	$(2\lambda, 6\mu, +)$, $(2\lambda+1, 6\mu, -)$
$6mm$	$6 \parallel z, m \parallel y$	$(l, 6\mu, +)$
$\bar{6}m2$	$\bar{6} \parallel z, m \perp y$	$(2\lambda, 6\mu, +)$, $(2\lambda+1, 6\mu+3, +)$
	$m \perp x$	$(2\lambda, 6\mu, +)$, $(2\lambda+1, 6\mu+3, -)$
$6/mmm$	$6 \parallel z, m \perp z, m \perp y$	$(2\lambda, 6\mu, +)$

The Fourier transform of the product of a complex spherical harmonic function with normalization $\int |Y_{lm}|^2 \, \mathrm{d}\Omega = 1$ and an arbitrary radial function $R_l(r)$ follows from the orthonormality properties of the spherical harmonic functions, and is given by

$$\int Y_{lm} R_l(r) \exp(2\pi i \mathbf{S} \cdot \mathbf{r}) \, \mathrm{d}\tau = 4\pi i^l \int j_l(2\pi S r) R_l(r) r^2 \, \mathrm{d}r Y_{lm}(\beta, \gamma), \tag{1.2.7.8a}$$

where j_l is the lth-order spherical Bessel function (Arfken, 1970), and θ and φ, β and γ are the angular coordinates of $\mathbf{r}$ and $\mathbf{S}$, respectively.

For the Fourier transform of the real spherical harmonic functions, the scattering operator is expressed in terms of the real spherical harmonics:

$$\exp(2\pi i \mathbf{S} \cdot \mathbf{r}) = \sum_{l=0}^{\infty} i^l j_l(2\pi S r)(2 - \delta_{m0})(2l+1) \sum_{m=0}^{l} \frac{(l-m)!}{(l+m)!} \times P_l^m(\cos\theta) P_l^m(\cos\beta) \cos[m(\phi - \gamma)], \tag{1.2.7.7b}$$

which leads to

$$\int y_{lmp}(\theta, \varphi) R_l(r) \exp(2\pi i \mathbf{S} \cdot \mathbf{r}) \, \mathrm{d}\tau = 4\pi i^l \langle j_l \rangle y_{lmp}(\beta, \gamma). \tag{1.2.7.8b}$$

Since y_{lmp} occurs on both sides, the expression is independent of the normalization selected. Therefore, for the Fourier transform of the density functions d_{lmp}

$$\int d_{lmp}(\theta, \varphi) R_l(r) \exp(2\pi i \mathbf{S} \cdot \mathbf{r}) \, \mathrm{d}\tau = 4\pi i^l \langle j_l \rangle d_{lmp}(\beta, \gamma). \tag{1.2.7.8c}$$

In (1.2.7.8*b*) and (1.2.7.8*c*), $\langle j_l \rangle$, the *Fourier–Bessel transform*, is the radial integral defined as

$$\langle j_l \rangle = \int j_l(2\pi S r) R_l(r) r^2 \, \mathrm{d}r \tag{1.2.7.9}$$

of which $\langle j_0 \rangle$ in expression (1.2.4.3) is a special case. The functions $\langle j_l \rangle$ for Hartree–Fock valence shells of the atoms are tabulated in scattering-factor tables (*IT* IV, 1974). Expressions for the evaluation of $\langle j_l \rangle$ using the radial function (1.2.7.5*a–c*) have been given by Stewart (1980) and in closed form for (1.2.7.5*a*) by Avery & Watson (1977) and Su & Coppens (1990). The closed-form expressions are listed in Table 1.2.7.4.

Expressions (1.2.7.8) show that the Fourier transform of a direct-space spherical harmonic function is a reciprocal-space spherical harmonic function with the same l, m, or, in other words, the spherical harmonic functions are Fourier-transform invariant.

The scattering factors $f_{lmp}(\mathbf{S})$ of the aspherical density functions $R_l(r) d_{lmp}(\theta, \phi)$ in the multipole expansion (1.2.7.6) are thus given by

$$f_{lmp}(\mathbf{S}) = 4\pi i^l \langle j_l \rangle d_{lmp}(\beta, \gamma). \tag{1.2.7.8d}$$

The reciprocal-space spherical harmonic functions in this expression are identical to the functions given in Table 1.2.7.1, except for the replacement of the direction cosines x, y and z by the direction cosines of the scattering vector $\mathbf{S}$.

1.2.8. Fourier transform of orbital products

If the wavefunction is written as a sum over normalized Slater determinants, each representing an antisymmetrized combination of occupied molecular orbitals χ_i expressed as linear combinations of atomic orbitals φ_ν, *i.e.* $\chi_i = \sum_\nu c_{i\nu} \varphi_\nu$, the electron density is given by (Stewart, 1969*a*)

$$\rho(\mathbf{r}) = \sum_i n_i \chi_i^2 = \sum_\mu \sum_\nu P_{\mu\nu} \varphi_\mu(\mathbf{r}) \varphi_\nu(\mathbf{r}), \tag{1.2.8.1}$$

with $n_i = 1$ or 2. The coefficients $P_{\mu\nu}$ are the populations of the orbital product density functions $\phi_\mu(\mathbf{r})\varphi_\nu(\mathbf{r})$ and are given by

$$P_{\mu\nu} = \sum_i n_i c_{i\mu} c_{i\nu}. \tag{1.2.8.2}$$

For a multi-Slater determinant wavefunction the electron density is expressed in terms of the occupied natural spin orbitals, leading again to (1.2.8.2) but with noninteger values for the coefficients n_i.

The summation (1.2.8.1) consists of one- and two-centre terms for which φ_μ and φ_ν are centred on the same or on different nuclei, respectively. The latter represent the overlap density, which is only significant if $\varphi_\mu(\mathbf{r})$ and $\varphi_\nu(\mathbf{r})$ have an appreciable value in the same region of space.

1.2.8.1. One-centre orbital products

If the atomic basis consists of hydrogenic type s, p, d, f, ... orbitals, the basis functions may be written as

$$\varphi(r, \theta, \varphi) = R_l(r) Y_{lm}(\theta, \varphi) \tag{1.2.8.3a}$$

or

$$\varphi(r, \theta, \varphi) = R_l(r) y_{lmp}(\theta, \varphi), \tag{1.2.8.3b}$$

which gives for corresponding values of the orbital products

$$\varphi_\mu(\mathbf{r}) \varphi_\nu(\mathbf{r}) = R_l(r) R_{l'}(r) Y_{lm}(\theta, \varphi) Y_{l'm'}(\theta, \varphi) \tag{1.2.8.4a}$$

and

$$\varphi_\mu(\mathbf{r}) \varphi_\nu(\mathbf{r}) = R_l(r) R_{l'}(r) y_{lmp}(\theta, \varphi) y_{l'm'p'}(\theta, \varphi), \tag{1.2.8.4b}$$

Table 1.2.7.3. *'Kubic harmonic' functions*

(*a*) Coefficients in the expression $K_{lj} = \sum_{mp} k^{l}_{mpj} y_{lmp}$ with normalization $\int_0^{\pi}\int_0^{2\pi} |K_{lj}|^2 \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi = 1$ (Kara & Kurki-Suonio, 1981).

Even l		mp					
l	j	0+	2+	4+	6+	8+	10+
0	1	1					
4	1	$\frac{1}{2}\left(\frac{7}{3}\right)^{1/2}$ 0.76376		$\frac{1}{2}\left(\frac{5}{3}\right)^{1/2}$ 0.64550			
6	1	$\frac{1}{2}\left(\frac{1}{2}\right)^{1/2}$ 0.35355		$-\frac{1}{2}\left(\frac{7}{2}\right)^{1/2}$ −0.93541			
6	2		$\frac{1}{4}11^{1/2}$ 0.82916		$-\frac{1}{4}5^{1/2}$ −0.55902		
8	1	$\frac{1}{8}33^{1/2}$ 0.71807		$\frac{1}{4}\left(\frac{7}{3}\right)^{1/2}$ 0.38188		$\frac{1}{8}\left(\frac{65}{3}\right)^{1/2}$ 0.58184	
10	1	$\frac{1}{8}\left(\frac{65}{6}\right)^{1/2}$ 0.41143		$-\frac{1}{4}\left(\frac{11}{2}\right)^{1/2}$ −0.58630		$-\frac{1}{8}\left(\frac{187}{6}\right)^{1/2}$ −0.69784	
10	2		$\frac{1}{8}\left(\frac{247}{6}\right)^{1/2}$ 0.80202		$\frac{1}{16}\left(\frac{19}{3}\right)^{1/2}$ 0.15729		$\frac{1}{16}85^{1/2}$ 0.57622
l	j		2−	4−	6−	8−	
3	1		1				
7	1		$\frac{1}{2}\left(\frac{13}{6}\right)^{1/2}$ 0.73598		$\frac{1}{2}\left(\frac{11}{16}\right)^{1/2}$ 0.41458		
9	1		$\frac{1}{4}3^{1/2}$ 0.43301		$-\frac{1}{4}13^{1/2}$ −0.90139		
9	2		$\frac{1}{2}\left(\frac{17}{6}\right)^{1/2}$ 0.84163		$-\frac{1}{2}\left(\frac{7}{6}\right)^{1/2}$ −0.54006		

(*b*) Coefficients k^{l}_{mpj} and density normalization factors N_{lj} in the expression $K_{lj} = N_{lj}\sum_{mp} k^{l}_{mpj} u_{lmp}$ where $u_{lm\pm} = P^{m}_{l}(\cos\theta)^{\cos m\varphi}_{\sin m\varphi}$ (Su & Coppens, 1994).

Even l		N_{lj}	mp					
l	j		0+	2+	4+	6+	8+	10+
0	1	$1/4\pi = 0.079577$	1					
4	1	0.43454	1		+1/168			
6	1	0.25220	1		−1/360			
6	2	0.020833		1		−1/792		
8	1	0.56292	1		1/5940		$\frac{1}{672}\times\frac{1}{5940}$	
10	1	0.36490	1		1/5460		$\frac{1}{4320}\times\frac{1}{5460}$	
10	2	0.0095165	1			1/43680		$-\frac{1}{456}\times\frac{1}{43680}$
l	j			2−	4−	6−	8−	
3	1	0.066667		1				
7	1	0.014612		1		1/1560		
9	1	0.0059569		1		1/2520		
9	2	0.00014800			1		−1/4080	

(*c*) Density-normalized *Kubic harmonics* as linear combinations of density-normalized spherical harmonic functions. Coefficients in the expression $K_{lj} = \sum_{mp} k''^{l}_{mpj} d_{lmp}$. Density-type normalization is defined as $\int_0^{\pi}\int_0^{2\pi} |K_{lj}| \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi = 2 - \delta_{l0}$.

Even l		mp					
l	j	0+	2+	4+	6+	8+	10+
0	1	1					
4	1	0.78245		0.57939			
6	1	0.37790		−0.91682			
6	2		0.83848		−0.50000		
l	j	2−	4−	6−	8−		
3	1	1					
7	1	0.73145		0.63290			

Table 1.2.7.3 (*cont.*)

(*d*) Index rules for cubic symmetries (Kurki-Suonio, 1977; Kara & Kurki-Suonio, 1981).

l	j	23 T	$m\bar{3}$ T_h	432 O	$\bar{4}3m$ T_d	$m\bar{3}m$ O_h
0	1	×	×	×	×	×
3	1	×			×	
4	1	×	×	×	×	×
6	1	×	×	×	×	×
6	2	×	×			
7	1	×			×	
8	1	×	×	×	×	×
9	1	×			×	
9	2	×		×		
10	1	×	×	×	×	×
10	2	×	×			

respectively, where it has been assumed that the radial function depends only on l.

Because the spherical harmonic functions form a complete set, their products can be expressed as a linear combination of spherical harmonics. The coefficients in this expansion are the *Clebsch–Gordan coefficients* (Condon & Shortley, 1957), defined by

$$Y_{lm}(\theta,\varphi)Y_{l'm'}(\theta,\varphi)=\sum_L\sum_M C_{Ll l'}^{Mmm'}\,Y_{LM}(\theta,\varphi) \qquad (1.2.8.5a)$$

or the equivalent definition

$$C_{Ll l'}^{Mmm'}=\int_0^{\pi}\sin\theta\,\mathrm{d}\theta\int_0^{2\pi}\mathrm{d}\varphi Y_{LM}^{*}(\theta,\varphi)Y_{lm}(\theta,\varphi)Y_{l'm'}(\theta,\varphi). \quad (1.2.8.5b)$$

The $C_{Ll l'}^{Mmm'}$ vanish, unless $L+l+l'$ is even, $|l-l'|<L<l+l'$ and $M=m+m'$.

The corresponding expression for y_{lmp} is

$$y_{lmp}(\theta,\varphi)y_{l'm'p'}(\theta,\varphi)=\sum_L\sum_M C'^{\,Mmm'}_{\substack{Ll l'\\ P}}\,y_{LMP}(\theta,\varphi), \qquad (1.2.8.5c)$$

with $M=|m+m'|$ and $|m-m'|$ for $p=p'$, and $M=-|m+m'|$ and $-|m-m'|$ for $p=-p'$ and $P=p\times p'$.

Values of C and C' for $l\le 2$ are given in Tables 1.2.8.1 and 1.2.8.2. They are valid for the functions Y_{lm} and y_{lmp} with normalization $\int|Y_{lm}|^2\,\mathrm{d}\Omega=1$ and $\int y_{lmp}^2\,\mathrm{d}\Omega=1$.

By using (1.2.8.5*a*) or (1.2.8.5*c*), the one-centre orbital products are expressed as a sum of spherical harmonic functions. It follows that the one-centre orbital product density basis set is formally equivalent to the multipole description, both in real and in reciprocal space. To obtain the relation between orbital products and the charge-density functions, the right-hand side of (1.2.8.5*c*) has to be multiplied by the ratio of the normalization constants, as the wavefunctions y_{lmp} and charge-density functions d_{lmp} are normalized in a different way as described by (1.2.7.3*a*) and (1.2.7.3*b*). Thus

$$y_{lmp}(\theta,\varphi)y_{l'm'p'}(\theta,\varphi)=\sum_L\sum_M R_{LMP}C'^{\,Mmm'}_{\substack{Ll l'\\ P}}\,d_{LMP}(\theta,\varphi), \quad (1.2.8.6)$$

where $R_{LMP}=M_{LMP}$ (wavefunction)$/L_{LMP}$ (density function). The normalization constants M_{lmp} and L_{lmp} are given in Table 1.2.7.1, while the coefficients in the expressions (1.2.8.6) are listed in Table 1.2.8.3.

1.2.8.2. Two-centre orbital products

Fourier transform of the electron density as described by (1.2.8.1) requires explicit expressions for the two-centre orbital product scattering. Such expressions are described in the literature for both Gaussian (Stewart, 1969*b*) and Slater-type (Bentley & Stewart, 1973; Avery & Ørmen, 1979) atomic orbitals. The expressions can also be used for Hartree–Fock atomic functions, as expansions in terms of Gaussian- (Stewart, 1969*b*, 1970; Stewart & Hehre, 1970; Hehre *et al.*, 1970) and Slater-type (Clementi & Roetti, 1974) functions are available for many atoms.

1.2.9. The atomic temperature factor

Since the crystal is subject to vibrational oscillations, the observed elastic scattering intensity is an average over all normal modes of the crystal. Within the Born–Oppenheimer approximation, the theoretical electron density should be calculated for each set of nuclear coordinates. An average can be obtained by taking into account the statistical weight of each nuclear configuration, which may be expressed by the probability distribution function $P(\mathbf{u}_1,\ldots,\mathbf{u}_N)$ for a set of displacement coordinates $\mathbf{u}_1,\ldots,\mathbf{u}_N$.

In general, if $\rho(\mathbf{r},\mathbf{u}_1,\ldots,\mathbf{u}_N)$ is the electron density corresponding to the geometry defined by $\mathbf{u}_1,\ldots,\mathbf{u}_N$, the time-averaged electron density is given by

$$\langle\rho(\mathbf{r})\rangle=\int\rho(\mathbf{r},\mathbf{u}_1,\ldots,\mathbf{u}_N)P(\mathbf{u}_1,\ldots,\mathbf{u}_N)\,\mathrm{d}\mathbf{u}_1\ldots\mathrm{d}\mathbf{u}_N. \quad (1.2.9.1)$$

When the crystal can be considered as consisting of perfectly following rigid entities, which may be molecules or atoms, expression (1.2.9.1) simplifies:

$$\langle\rho_{\text{rigid group}}(\mathbf{r})\rangle=\int\rho_{\text{r.g., static}}(\mathbf{r}-\mathbf{u})P(\mathbf{u})\,\mathrm{d}\mathbf{u}=\rho_{\text{r.g., static}}*P(\mathbf{u}). \quad (1.2.9.2)$$

In the approximation that the atomic electrons perfectly follow the nuclear motion, one obtains

$$\langle\rho_{\text{atom}}(\mathbf{r})\rangle=\rho_{\text{atom, static}}(\mathbf{r})*P(\mathbf{u}). \quad (1.2.9.3)$$

The Fourier transform of this convolution is the product of the Fourier transforms of the individual functions:

$$\langle f(\mathbf{H})\rangle=f(\mathbf{H})T(\mathbf{H}). \quad (1.2.9.4)$$

Thus $T(\mathbf{H})$, the atomic temperature factor, is the Fourier transform of the probability distribution $P(\mathbf{u})$.

1.2.10. The vibrational probability distribution and its Fourier transform in the harmonic approximation

For a harmonic oscillator, the probability distribution averaged over all populated energy levels is a Gaussian, centred at the equilibrium position. For the three-dimensional isotropic harmonic oscillator, the distribution is

Table 1.2.7.4. *Closed-form expressions for Fourier transform of Slater-type functions (Avery & Watson, 1977; Su & Coppens, 1990)*

$\langle j_k \rangle \equiv \int_0^\infty r^N \exp(-Zr) j_k(Kr)\,\mathrm{d}r$, $K = 4\pi \sin\theta/\lambda$.

	N							
k	1	2	3	4	5	6	7	8
0	$\frac{1}{K^2+Z^2}$	$\frac{2Z}{(K^2+Z^2)^2}$	$\frac{2(3Z^2-K^2)}{(K^2+Z^2)^3}$	$\frac{24Z(Z^2-K^2)}{(K^2+Z^2)^4}$	$\frac{24(5Z^2-10K^2Z^2+K^4)}{(K^2+Z^2)^5}$	$\frac{240Z(K^2-3Z^2)(3K^2-Z^2)}{(K^2+Z^2)^6}$	$\frac{720(7Z^6-35K^2Z^4+21K^4Z^2-K^6)}{(K^2+Z^2)^7}$	$\frac{40320(Z^7-7K^2Z^5+7K^4Z^3-K^6Z)}{(K^2+Z^2)^8}$
1		$\frac{2K}{(K^2+Z^2)^2}$	$\frac{8KZ}{(K^2+Z^2)^3}$	$\frac{8K(5Z^2-K^2)}{(K^2+Z^2)^4}$	$\frac{48KZ(5Z^2-3K^2)}{(K^2+Z^2)^5}$	$\frac{48K(35Z^4-42K^2Z^2+3K^4)}{(K^2+Z^2)^6}$	$\frac{1920KZ(7Z^4-14K^2Z^2+3K^4)}{(K^2+Z^2)^7}$	$\frac{5760K(21Z^6-63K^2Z^4+27K^4Z^2-K^6)}{(K^2+Z^2)^8}$
2			$\frac{8K^2}{(K^2+Z^2)^3}$	$\frac{48K^2Z}{(K^2+Z^2)^4}$	$\frac{48K^2(7Z^2-K^2)}{(K^2+Z^2)^5}$	$\frac{384K^2Z(7Z^2-3K^2)}{(K^2+Z^2)^6}$	$\frac{1152K^2(21Z^4-18K^2Z^2+K^4)}{(K^2+Z^2)^7}$	$\frac{11520K^2Z(21Z^4-30K^2Z^2+5K^4)}{(K^2+Z^2)^8}$
3				$\frac{48K^3}{(K^2+Z^2)^4}$	$\frac{384K^3Z}{(K^2+Z^2)^5}$	$\frac{384K^3(9Z^2-K^2)}{(K^2+Z^2)^6}$	$\frac{11520K^3Z(3Z^2-K^2)}{(K^2+Z^2)^7}$	$\frac{11520K^3(33Z^4-22K^2Z^2+K^4)}{(K^2+Z^2)^8}$
4					$\frac{384K^4}{(K^2+Z^2)^5}$	$\frac{3840K^4Z}{(K^2+Z^2)^6}$	$\frac{3840K^4(11Z^2-K^2)}{(K^2+Z^2)^7}$	$\frac{46080K^4Z(11Z^2-3K^2)}{(K^2+Z^2)^8}$
5						$\frac{3840K^5}{(K^2+Z^2)^6}$	$\frac{46080K^5Z}{(K^2+Z^2)^7}$	$\frac{40680K^5(13Z^2-K^2)}{(K^2+Z^2)^8}$
6							$\frac{46080K^6}{(K^2+Z^2)^7}$	$\frac{645120K^6Z}{(K^2+Z^2)^8}$
7								$\frac{645120K^7}{(K^2+Z^2)^8}$

$$P(u) = (2\pi\langle u^2\rangle)^{-3/2} \exp\{-|u|^2/2\langle u^2\rangle\}, \tag{1.2.10.1}$$

where $\langle u^2\rangle$ is the mean-square displacement in any direction.

The corresponding trivariate normal distribution to be used for anisotropic harmonic motion is, in tensor notation,

$$P(\mathbf{u}) = \frac{|\boldsymbol{\sigma}^{-1}|^{1/2}}{(2\pi)^{3/2}} \exp\{-\tfrac{1}{2}\boldsymbol{\sigma}^{-1}_{jk}(u^j u^k)\}. \tag{1.2.10.2a}$$

Here $\boldsymbol{\sigma}$ is the variance–covariance matrix, with covariant components, and $|\boldsymbol{\sigma}^{-1}|$ is the determinant of the inverse of $\boldsymbol{\sigma}$. Summation over repeated indices has been assumed. The corresponding equation in matrix notation is

$$P(\mathbf{u}) = \frac{|\boldsymbol{\sigma}^{-1}|^{1/2}}{(2\pi)^{3/2}} \exp\{-\tfrac{1}{2}(\mathbf{u})^T \boldsymbol{\sigma}^{-1}(\mathbf{u})\}, \tag{1.2.10.2b}$$

where the superscript T indicates the transpose.

The characteristic function, or Fourier transform, of $P(\mathbf{u})$ is

$$T(\mathbf{H}) = \exp\{-2\pi^2 \sigma^{jk} h_j h_k\} \tag{1.2.10.3a}$$

or

$$T(\mathbf{H}) = \exp\{-2\pi^2 \mathbf{H}^T \boldsymbol{\sigma} \mathbf{H}\}. \tag{1.2.10.3b}$$

With the change of variable $b^{jk} = 2\pi^2\sigma^{jk}$, (1.2.10.3*a*) becomes

$$T(\mathbf{H}) = \exp\{-b^{jk} h_j h_k\}.$$

1.2.11. Rigid-body analysis

The treatment of rigid-body motion of molecules or molecular fragments was developed by Cruickshank (1956) and expanded into a general theory by Schomaker & Trueblood (1968). The theory has been described by Johnson (1970*b*) and by Dunitz (1979). The latter reference forms the basis for the following treatment.

The most general motions of a rigid body consist of rotations about three axes, coupled with translations parallel to each of the axes. Such motions correspond to screw rotations. A libration around a vector $\boldsymbol{\lambda}$ (λ_1, λ_2, λ_3), with length corresponding to the magnitude of the rotation, results in a displacement $\delta\mathbf{r}$, such that

$$\delta\mathbf{r} = (\boldsymbol{\lambda} \times \mathbf{r}) = \mathbf{D}\mathbf{r} \tag{1.2.11.1}$$

with

$$\mathbf{D} = \begin{bmatrix} 0 & -\lambda_3 & \lambda_2 \\ \lambda_3 & 0 & -\lambda_1 \\ -\lambda_2 & \lambda_1 & 0 \end{bmatrix}, \tag{1.2.11.2}$$

or in Cartesian tensor notation, assuming summation over repeated indices,

$$\delta r_i = D_{ij} r_j = -\varepsilon_{ijk} \lambda_k r_j \tag{1.2.11.3}$$

where the permutation operator ε_{ijk} equals +1 for i, j, k a cyclic permutation of the indices 1, 2, 3, or −1 for a noncyclic permutation, and zero if two or more indices are equal. For i = 1, for example, only the ε_{123} and ε_{132} terms occur. Addition of a translational displacement gives

$$\delta r_i = D_{ij} r_j + t_i. \tag{1.2.11.4}$$

When a rigid body undergoes vibrations the displacements vary with time, so suitable averages must be taken to derive the mean-square displacements. If the librational and translational motions are independent, the cross products between the two terms in (1.2.11.4) average to zero and the elements of the mean-square displacement tensor of atom n, U^n_{ij}, are given by

$$\begin{aligned}
U^n_{11} &= +L_{22}r_3^2 + L_{33}r_2^2 - 2L_{23}r_2r_3 + T_{11} \\
U^n_{22} &= +L_{33}r_1^2 + L_{11}r_3^2 - 2L_{13}r_1r_3 + T_{22} \\
U^n_{33} &= +L_{11}r_2^2 + L_{22}r_1^2 - 2L_{12}r_1r_2 + T_{33} \\
U^n_{12} &= -L_{33}r_1r_2 - L_{12}r_3^2 + L_{13}r_2r_3 + L_{23}r_1r_3 + T_{12} \\
U^n_{13} &= -L_{22}r_1r_3 + L_{12}r_2r_3 - L_{13}r_2^2 + L_{23}r_1r_2 + T_{13} \\
U^n_{23} &= -L_{11}r_2r_3 + L_{12}r_1r_3 - L_{13}r_1r_2 - L_{23}r_1^2 + T_{23},
\end{aligned} \tag{1.2.11.5}$$

where the coefficients $L_{ij} = \langle\lambda_i\lambda_j\rangle$ and $T_{ij} = \langle t_i t_j\rangle$ are the elements of the 3 × 3 libration tensor $\mathbf{L}$ and the 3 × 3 translation tensor $\mathbf{T}$, respectively. Since pairs of terms such as $\langle t_i t_j\rangle$ and $\langle t_j t_i\rangle$

Table 1.2.8.1. *Products of complex spherical harmonics as defined by equation (1.2.7.2a)*

$$
\begin{aligned}
Y_{00}\ Y_{00} &= 0.28209479Y_{00}\\
Y_{10}\ Y_{00} &= 0.28209479Y_{10}\\
Y_{10}\ Y_{10} &= 0.25231325Y_{20} + 0.28209479Y_{00}\\
Y_{11}\ Y_{00} &= 0.28209479Y_{11}\\
Y_{11}\ Y_{10} &= 0.21850969Y_{21}\\
Y_{11}\ Y_{11} &= 0.30901936Y_{22}\\
Y_{11}\ Y_{11-} &= -0.12615663Y_{20} + 0.28209479Y_{00}\\
Y_{20}\ Y_{00} &= 0.28209479Y_{20}\\
Y_{20}\ Y_{10} &= 0.24776669Y_{30} + 0.25231325Y_{10}\\
Y_{20}\ Y_{11} &= 0.20230066Y_{31} - 0.12615663Y_{11}\\
Y_{20}\ Y_{20} &= 0.24179554Y_{40} + 0.18022375Y_{20} + 0.28209479Y_{00}\\
Y_{21}\ Y_{00} &= 0.28209479Y_{21}\\
Y_{21}\ Y_{10} &= 0.23359668Y_{31} + 0.21850969Y_{11}\\
Y_{21}\ Y_{11} &= 0.26116903Y_{32}\\
Y_{21}\ Y_{11-} &= -0.14304817Y_{30} + 0.21850969Y_{10}\\
Y_{21}\ Y_{20} &= 0.22072812Y_{41} + 0.09011188Y_{21}\\
Y_{21}\ Y_{21} &= 0.25489487Y_{42} + 0.22072812Y_{22}\\
Y_{21}\ Y_{21-} &= -0.16119702Y_{40} + 0.09011188Y_{20} + 0.28209479Y_{00}\\
Y_{22}\ Y_{00} &= 0.28209479Y_{22}\\
Y_{22}\ Y_{10} &= 0.18467439Y_{32}\\
Y_{22}\ Y_{11} &= 0.31986543Y_{33}\\
Y_{22}\ Y_{11-} &= -0.08258890Y_{31} + 0.30901936Y_{11}\\
Y_{22}\ Y_{20} &= 0.15607835Y_{42} - 0.18022375Y_{22}\\
Y_{22}\ Y_{21} &= 0.23841361Y_{43}\\
Y_{22}\ Y_{21-} &= -0.09011188Y_{41} + 0.22072812Y_{21}\\
Y_{22}\ Y_{22} &= 0.33716777Y_{44}\\
Y_{22}\ Y_{22-} &= 0.04029926Y_{40} - 0.18022375Y_{20} + 0.28209479Y_{00}
\end{aligned}
$$

Table 1.2.8.2. *Products of real spherical harmonics as defined by equations (1.2.7.2b) and (1.2.7.2c)*

$$
\begin{aligned}
y_{00}\ y_{00} &= 0.28209479y_{00}\\
y_{10}\ y_{00} &= 0.28209479y_{10}\\
y_{10}\ y_{10} &= 0.25231325y_{20} + 0.28209479y_{00}\\
y_{11\pm}\ y_{00} &= 0.28209479y_{11\pm}\\
y_{11\pm}\ y_{10} &= 0.21850969y_{21\pm}\\
y_{11\pm}\ y_{11\pm} &= 0.21850969y_{22+} - 0.12615663y_{20} + 0.28209479y_{00}\\
y_{11+}\ y_{11-} &= 0.21850969y_{22-}\\
y_{20}\ y_{00} &= 0.28209479y_{20}\\
y_{20}\ y_{10} &= 0.24776669y_{30} + 0.25231325y_{10}\\
y_{20}\ y_{11\pm} &= 0.20230066y_{31\pm} - 0.12615663y_{11\pm}\\
y_{20}\ y_{20} &= 0.24179554y_{40} + 0.18022375y_{20} + 0.28209479y_{00}\\
y_{21\pm}\ y_{00} &= 0.28209479y_{21\pm}\\
y_{21\pm}\ y_{10} &= 0.23359668y_{31\pm} + 0.21850969y_{11\pm}\\
y_{21\pm}\ y_{11\pm} &= \pm 0.18467439y_{32+} - 0.14304817y_{30} + 0.21850969y_{10}\\
y_{21\pm}\ y_{11\mp} &= 0.18467469y_{32-}\\
y_{21\pm}\ y_{20} &= 0.22072812y_{41\pm} + 0.09011188y_{21\pm}\\
y_{21\pm}\ y_{21\pm} &= \pm 0.18022375y_{42+} \pm 0.15607835y_{22+}\\
&\quad - 0.16119702y_{40} + 0.09011188y_{20} + 0.28209479y_{00}\\
y_{21+}\ y_{21-} &= -0.18022375y_{42-} + 0.15607835y_{22-}\\
y_{22\pm}\ y_{00} &= 0.28209479y_{22\pm}\\
y_{22\pm}\ y_{10} &= 0.18467439y_{32\pm}\\
y_{22\pm}\ y_{11\pm} &= \pm 0.22617901y_{33+} - 0.05839917y_{31+} + 0.21850969y_{11+}\\
y_{22\pm}\ y_{11\mp} &= 0.22617901y_{33-} \pm 0.05839917y_{31-} \mp 0.21850969y_{11-}\\
y_{22\pm}\ y_{20} &= 0.15607835y_{42\pm} - 0.18022375y_{22\pm}\\
y_{22\pm}\ y_{21\pm} &= \pm 0.16858388y_{43+} - 0.06371872y_{41+} + 0.15607835y_{21+}\\
y_{22\pm}\ y_{21\mp} &= 0.16858388y_{43-} \pm 0.06371872y_{41-} \mp 0.15607835y_{21-}\\
y_{22\pm}\ y_{22\pm} &= \pm 0.23841361y_{44+} + 0.04029926y_{40} - 0.18022375y_{20}\\
&\quad + 0.28209479y_{00}\\
y_{22+}\ y_{22-} &= 0.23841361y_{44-}
\end{aligned}
$$

correspond to averages over the same two scalar quantities, the **T** and **L** tensors are symmetrical.

If a rotation axis is correctly oriented, but incorrectly positioned, an additional translation component perpendicular to the rotation axes is introduced. The rotation angle and the parallel component of the translation are invariant to the position of the axis, but the perpendicular component is not. This implies that the **L** tensor is unaffected by any assumptions about the position of the libration axes, whereas the **T** tensor depends on the assumptions made concerning the location of the axes.

The quadratic correlation between librational and translational motions can be allowed for by including in (1.2.11.5) cross terms of the type $\langle D_{ik}t_j\rangle$, or, with (1.2.11.3),

$$
\begin{aligned}
U_{ij} &= \langle D_{ik}D_{jl}\rangle r_k r_l + \langle D_{ik}t_j + D_{jk}t_i\rangle r_k + \langle t_i t_j\rangle\\
&= A_{ijkl}r_k r_l + B_{ijk}r_k + \langle t_i t_j\rangle,
\end{aligned}
\tag{1.2.11.6}
$$

which leads to the explicit expressions such as

$$
\begin{aligned}
U_{11} = \langle \delta r_1\rangle^2 &= \langle\lambda_3^2\rangle r_2^2 + \langle\lambda_2^2\rangle r_3^2 - 2\langle\lambda_2\lambda_3\rangle r_2 r_3\\
&\quad - 2\langle\lambda_3 t_1\rangle r_2 - 2\langle\lambda_2 t_1\rangle r_3 + \langle t_1^2\rangle,\\
U_{12} = \langle \delta r_1 \delta r_2\rangle &= -\langle\lambda_3^2\rangle r_1 r_2 + \langle\lambda_1\lambda_3\rangle r_2 r_3 + \langle\lambda_2\lambda_3\rangle r_1 r_3\\
&\quad - \langle\lambda_1\lambda_2\rangle r_3^2 + \langle\lambda_3 t_1\rangle r_1 - \langle\lambda_1 t_1\rangle r_3\\
&\quad - \langle\lambda_3 t_2\rangle r_2 + \langle\lambda_2 t_2\rangle r_3 + \langle t_1 t_2\rangle.
\end{aligned}
\tag{1.2.11.7}
$$

The products of the type $\langle\lambda_i t_j\rangle$ are the components of an additional tensor, **S**, which unlike the tensors **T** and **L** is unsymmetrical, since $\langle\lambda_i t_j\rangle$ is different from $\langle\lambda_j t_i\rangle$. The terms involving elements of **S** may be grouped as

$$
\langle\lambda_3 t_1\rangle r_1 - \langle\lambda_3 t_2\rangle r_2 + (\langle\lambda_2 t_2\rangle - \langle\lambda_1 t_1\rangle) r_3 \tag{1.2.11.8}
$$

or

$$
S_{31}r_1 - S_{32}r_2 + (S_{22} - S_{11})r_3.
$$

As the diagonal elements occur as differences in this expression, a constant may be added to each of the diagonal terms without changing the observational equations. In other words, the trace of **S** is indeterminate.

In terms of the **L**, **T** and **S** tensors, the observational equations are

$$
U_{ij} = G_{ijkl}L_{kl} + H_{ijkl}S_{kl} + T_{ij}. \tag{1.2.11.9}
$$

The arrays G_{ijkl} and H_{ijkl} involve the atomic coordinates $(x,\ y,\ z) = (r_1,\ r_2,\ r_3)$, and are listed in Table 1.2.11.1. Equations (1.2.11.9) for each of the atoms in the rigid body form the observational equations, from which the elements of **T**, **L** and **S** can be derived by a linear least-squares procedure. One of the diagonal elements of **S** must be fixed in advance or some other suitable constraint applied because of the indeterminacy of Tr(**S**). It is common practice to set Tr(**S**) equal to zero. There are thus eight elements of **S** to be determined, as well as the six each of **L** and **T**, for a total of 20 variables. A shift of origin leaves **L** invariant, but it intermixes **T** and **S**.

If the origin is located at a centre of symmetry, for each atom at **r** with vibration tensor $\mathbf{U}^n$ there will be an equivalent atom at $-\mathbf{r}$ with the same vibration tensor. When the observational equations for these two atoms are added, the terms involving elements of **S** disappear since they are linear in the components of **r**. The other terms, involving elements of the **T** and **L** tensors, are simply doubled, like the $\mathbf{U}^n$ components.

The physical meaning of the **T** and **L** tensor elements is as follows. $T_{ij}l_i l_j$ is the mean-square amplitude of translational vibration in the direction of the unit vector **l** with components $l_1,\ l_2,\ l_3$ along the Cartesian axes and $L_{ij}l_i l_j$ is the mean-square amplitude of libration about an axis in this direction. The quantity $S_{ij}l_i l_j$ represents the mean correlation between libration

Table 1.2.8.3. *Products of two real spherical harmonic functions y_{lmp} in terms of the density functions d_{lmp} defined by equation (1.2.7.3b)*

$$
\begin{aligned}
y_{00}\, y_{00} &= 1.0000 d_{00}\\
y_{10}\, y_{00} &= 0.43301 d_{10}\\
y_{10}\, y_{10} &= 0.38490 d_{20} + 1.0 d_{00}\\
y_{11\pm}\, y_{00} &= 0.43302 d_{11\pm}\\
y_{11\pm}\, y_{10} &= 0.31831 d_{21\pm}\\
y_{11\pm}\, y_{11\pm} &= 0.31831 d_{22+} - 0.19425 d_{20} + 1.0 d_{00}\\
y_{11+}\, y_{11-} &= 0.31831 d_{22-}\\
y_{20}\, y_{00} &= 0.43033 d_{20}\\
y_{20}\, y_{10} &= 0.37762 d_{30} + 0.38730 d_{10}\\
y_{20}\, y_{11\pm} &= 0.28864 d_{31\pm} - 0.19365 d_{11\pm}\\
y_{20}\, y_{20} &= 0.36848 d_{40} + 0.27493 d_{20} + 1.0 d_{00}\\
y_{21\pm}\, y_{00} &= 0.41094 d_{21\pm}\\
y_{21\pm}\, y_{10} &= 0.33329 d_{31\pm} + 0.33541 d_{11\pm}\\
y_{21\pm}\, y_{11\pm} &= \pm 0.26691 d_{32+} - 0.21802 d_{30} + 0.33541 d_{10}\\
y_{21\pm}\, y_{11\mp} &= -0.26691 d_{32-}\\
y_{21\pm}\, y_{20} &= 0.31155 d_{41\pm} + 0.13127 d_{21\pm}\\
y_{21\pm}\, y_{21\pm} &= \pm 0.25791 d_{42+} \pm 0.22736 d_{22+} - 0.24565 d_{40} + 0.13747 d_{20} + 1.0 d_{00}\\
y_{21+}\, y_{21-} &= 0.25790 d_{42-} + 0.22736 d_{22-}\\
y_{22\pm}\, y_{00} &= 0.41094 d_{22\pm}\\
y_{22\pm}\, y_{10} &= 0.26691 d_{32\pm}\\
y_{22\pm}\, y_{11\pm} &= \pm 0.31445 d_{33+} - 0.083323 d_{31+} + 0.33541 d_{11+}\\
y_{22\pm}\, y_{11\mp} &= 0.31445 d_{33-} \pm 0.083323 d_{31-} \mp 0.33541 d_{11-}\\
y_{22\pm}\, y_{20} &= 0.22335 d_{42\pm} - 0.26254 d_{22\pm}\\
y_{22\pm}\, y_{21\pm} &= \pm 0.23873 d_{43+} - 0.089938 d_{41+} + 0.22736 d_{21+}\\
y_{22\pm}\, y_{21\mp} &= 0.23873 d_{43-} \pm 0.089938 d_{41-} \mp 0.22736 d_{21-}\\
y_{22\pm}\, y_{22\pm} &= \pm 0.31831 d_{44+} + 0.061413 d_{40} - 0.27493 d_{20} + 1.0 d_{00}\\
y_{22+}\, y_{22-} &= 0.31831 d_{44-}
\end{aligned}
$$

about the axis l and translation parallel to this axis. This quantity, like $T_{ij}l_il_j$, depends on the choice of origin, although the sum of the two quantities is independent of the origin.

The nonsymmetrical tensor **S** can be written as the sum of a symmetric tensor with elements $S^S_{ij} = (S_{ij} + S_{ji})/2$ and a skew-symmetric tensor with elements $S^A_{ij} = (S_{ij} - S_{ji})/2$. Expressed in terms of principal axes, $\mathbf{S}^S$ consists of three principal screw correlations $\langle \lambda_l t_l \rangle$. Positive and negative screw correlations correspond to opposite senses of helicity. Since an arbitrary constant may be added to all three correlation terms, only the differences between them can be determined from the data.

The skew-symmetric part $\mathbf{S}^A$ is equivalent to a vector $(\boldsymbol{\lambda} \times \mathbf{t})/2$ with components $(\boldsymbol{\lambda} \times \mathbf{t})_i/2 = (\lambda_j t_k - \lambda_k t_j)/2$, involving correlations between a libration and a perpendicular translation. The components of $\mathbf{S}^A$ can be reduced to zero, and **S** made symmetric, by a change of origin. It can be shown that the origin shift that symmetrizes **S** also minimizes the trace of **T**. In terms of the coordinate system based on the principal axes of **L**, the required origin shifts $\widehat{\rho}_i$ are

$$\widehat{\rho}_1 = \frac{\widehat{S}_{23} - \widehat{S}_{32}}{\widehat{L}_{22} + \widehat{L}_{33}} \quad \widehat{\rho}_2 = \frac{\widehat{S}_{31} - \widehat{S}_{13}}{\widehat{L}_{11} + \widehat{L}_{33}} \quad \widehat{\rho}_3 = \frac{\widehat{S}_{12} - \widehat{S}_{21}}{\widehat{L}_{11} + \widehat{L}_{22}}, \tag{1.2.11.10}$$

in which the carets indicate quantities referred to the principal axis system.

The description of the averaged motion can be simplified further by shifting to three generally nonintersecting libration axes, one each for each principal axis of **L**. Shifts of the $\mathbf{L}_1$ axis in the $\mathbf{L}_2$ and $\mathbf{L}_3$ directions by

$${}^1\widehat{\rho}_2 = -\widehat{S}_{13}/\widehat{L}_{11} \text{ and } {}^1\widehat{\rho}_3 = \widehat{S}_{12}/\widehat{L}_{11}, \tag{1.2.11.11}$$

respectively, annihilate the S_{12} and S_{13} terms of the symmetrized **S** tensor and simultaneously effect a further reduction in Tr(**T**) (the presuperscript denotes the axis that is shifted, the subscript the direction of the shift component). Analogous equations for displacements of the $\mathbf{L}_2$ and $\mathbf{L}_3$ axes are obtained by permutation of the indices. If all three axes are appropriately displaced, only the diagonal terms of **S** remain. Referred to the principal axes of **L**, they represent screw correlations along these axes and are independent of origin shifts.

Table 1.2.11.1. *The arrays G_{ijkl} and H_{ijkl} to be used in the observational equations $U_{ij} = G_{ijkl}L_{kl} + H_{ijkl}S_{kl} + T_{ij}$ [equation (1.2.11.9)]*

G_{ijkl}

	kl					
ij	11	22	33	23	31	12
11	0	z^2	y^2	$-2yz$	0	0
22	z^2	0	x^2	0	$-2xz$	0
33	y^2	x^2	0	0	0	$-2xy$
23	$-yz$	0	0	$-x^2$	xy	xz
31	0	$-xz$	0	xy	$-y^2$	yz
12	0	0	$-xy$	xz	yz	$-z^2$

H_{ijkl}

	kl								
ij	11	22	33	23	31	12	32	13	21
11	0	0	0	0	$-2y$	0	0	0	$2z$
22	0	0	0	0	0	$-2z$	$2x$	0	0
33	0	0	0	$-2x$	0	0	0	$2y$	0
23	0	$-x$	x	0	0	y	0	$-z$	0
31	y	0	$-y$	z	0	0	0	0	$-x$
12	$-z$	z	0	0	x	0	$-y$	0	0

The elements of the reduced **T** are

$$\begin{aligned} {}^rT_{II} &= \widehat{T}_{II} - \sum_{K \neq I} (\widehat{S}_{KI})^2/\widehat{L}_{KK} \\ {}^rT_{IJ} &= \widehat{T}_{IJ} - \sum_K \widehat{S}_{KI}\widehat{S}_{KJ}/\widehat{L}_{KK}, \quad J \neq I. \end{aligned} \tag{1.2.11.12}$$

The resulting description of the average rigid-body motion is in terms of six independently distributed instantaneous motions – three screw librations about nonintersecting axes (with screw pitches given by $\widehat{S}_{11}/\widehat{L}_{11}$ *etc.*) and three translations. The parameter set consists of three libration and three translation amplitudes, six angles of orientation for the principal axes of **L** and **T**, six coordinates of axis displacement, and three screw pitches, one of which has to be chosen arbitrarily, again for a total of 20 variables.

Since diagonal elements of **S** enter into the expression for ${}^rT_{IJ}$, the indeterminacy of Tr(**S**) introduces a corresponding indeterminacy in ${}^r\mathbf{T}$. The constraint Tr(**S**) = 0 is unaffected by the various rotations and translations of the coordinate systems used in the course of the analysis.

1.2.12. Treatment of anharmonicity

The probability distribution (1.2.10.2) is valid in the case of rectilinear harmonic motion. If the deviations from Gaussian shape are not too large, distributions may be used which are expansions with the Gaussian distribution as the leading term. Three such distributions are discussed in the following sections.

1.2.12.1. The Gram–Charlier expansion

The three-dimensional Gram–Charlier expansion, introduced into thermal-motion treatment by Johnson & Levy (1974), is an expansion of a function in terms of the zero and higher derivatives of a normal distribution (Kendall & Stuart, 1958). If D_j is the operator $\mathrm{d}/\mathrm{d}u^j$,

Table 1.2.12.1. *Some Hermite polynomials (Johnson & Levy, 1974; Zucker & Schulz, 1982)*

$$
\begin{aligned}
H(\mathbf{u}) &= 1\\
H_j(\mathbf{u}) &= w_j\\
H_{jk}(\mathbf{u}) &= w_jw_k - p_{jk}\\
H_{jkl}(\mathbf{u}) &= w_jw_kw_l - (w_jp_{kl} + w_kp_{lj} + w_lp_{jk}) = w_jw_kw_l - 3w_{(j}p_{kl)}\\
H_{jklm}(\mathbf{u}) &= w_jw_kw_lw_m - 6w_{(j}w_kp_{lm)} + 3p_{(j}p_{lm)}\\
H_{jklmn}(\mathbf{u}) &= w_jw_kw_lw_mw_n - 10w_{(i}w_mw_np_{jk)} + 15w_{(n}p_{jk}p_{lm)}\\
H_{jklmnp}(\mathbf{u}) &= w_jw_kw_lw_mw_nw_p - 15w_{(j}w_kw_lw_mp_{jk)} + 45w_{(j}w_kp_{lm}p_{np)}\\
&\quad - 15p_{(j}p_{lm}p_{np)}
\end{aligned}
$$

where $w_j \equiv p_{jk}u^k$ and p_{jk} are the elements of σ^{-1}, defined in expression (1.2.10.2). Indices between brackets indicate that the term is to be averaged over all permutations which produce distinct terms, keeping in mind that $p_{jk} = p_{kj}$ and $w_jw_k = w_kw_j$ as illustrated for H_{jkl}.

$$P(\mathbf{u}) = [1 - c^jD_j + \frac{1}{2!}c^{jk}D_jD_k - \frac{1}{3!}c^{jkl}D_jD_kD_l + \ldots + (-1)^r\frac{c^{\alpha_1}\ldots c^{\alpha_r}}{r!}D_{\alpha_1}D_{\alpha_r}]P_0(\mathbf{u}), \quad (1.2.12.1)$$

where $P_0(\mathbf{u})$ is the harmonic distribution, $\alpha_i = 1, 2$ or 3, and the operator $D_{\alpha_1}\ldots D_{\alpha_r}$ is the rth partial derivative $\partial^r/(\partial u^{\alpha 1}\ldots\partial u^{\alpha r})$. Summation is again implied over repeated indices.

The differential operators D may be eliminated by the use of three-dimensional Hermite polynomials $H_{\alpha_1\ldots\alpha_r}$ defined, by analogy with the one-dimensional Hermite polynomials, by the expression

$$D_{\alpha_1}\ldots D_{\alpha_r}\exp(-\tfrac{1}{2}\sigma^{-1}_{jk}u^ju^k) = (-1)^rH_{\alpha_1\ldots\alpha_r}(\mathbf{u})\exp(-\tfrac{1}{2}\sigma^{-1}_{jk}u^ju^k), \quad (1.2.12.2)$$

which gives

$$P(\mathbf{u}) = \left[1 + \frac{1}{3!}c^{jkl}H_{jkl}(\mathbf{u}) + \frac{1}{4!}c^{jklm}H_{jklm}(\mathbf{u}) + \frac{1}{5!}c^{jklmn}H_{jklmn}(\mathbf{u}) + \frac{1}{6!}c^{jklmnp}H_{jklmnp}(\mathbf{u}) + \ldots\right]P_0(\mathbf{u}), \quad (1.2.12.3)$$

where the first and second terms have been omitted since they are equivalent to a shift of the mean and a modification of the harmonic term only. The permutations of $j, k, l\ldots$ here, and in the following sections, include all combinations which produce different terms.

The coefficients c, defined by (1.2.12.1) and (1.2.12.2), are known as the *quasimoments* of the frequency function $P(\mathbf{u})$ (Kutznetsov *et al.*, 1960). They are related in a simple manner to the moments of the function (Kendall & Stuart, 1958) and are invariant to permutation of indices. There are 10, 15, 21 and 28 components of c for orders 3, 4, 5 and 6, respectively. The multivariate Hermite polynomials are functions of the elements of σ^{-1}_{jk} and of u^k, and are given in Table 1.2.12.1 for orders ≤ 6 (*IT* IV, 1974; Zucker & Schulz, 1982).

The Fourier transform of (1.2.12.3) is given by

$$T(\mathbf{H}) = \left[1 - \frac{4}{3}\pi^3ic^{jkl}h_jh_kh_l + \frac{2}{3}\pi^4c^{jklm}h_jh_kh_lh_m + \frac{4}{15}\pi^5ic^{jklmn}h_jh_kh_lh_mh_n - \frac{4}{45}\pi^6c^{jklmnp}h_jh_kh_lh_mh_nh_p + \ldots\right]T_0(\mathbf{H}), \quad (1.2.12.4)$$

where $T_0(\mathbf{H})$ is the harmonic temperature factor. $T(\mathbf{H})$ is a power-series expansion about the harmonic temperature factor, with even and odd terms, respectively, real and imaginary.

1.2.12.2. The cumulant expansion

A second statistical expansion which has been used to describe the atomic probability distribution is that of Edgeworth (Kendall & Stuart, 1958; Johnson, 1969). It expresses the function $P(\mathbf{u})$ as

$$P(\mathbf{u}) = \exp\left(\kappa^jD_j + \frac{1}{2!}\kappa^{jk}D_jD_k - \frac{1}{3!}\kappa^{jkl}D_jD_kD_l + \frac{1}{4!}\kappa^{jklm}D_jD_kD_lD_m - \ldots\right)P_0(\mathbf{u}). \quad (1.2.12.5a)$$

Like the moments μ of a distribution, the cumulants κ are descriptive constants. They are related to each other (in the one-dimensional case) by the identity

$$\exp\left\{\kappa_1t + \frac{\kappa_2t^2}{2!} + \ldots\frac{\kappa_rt^r}{r!} + \ldots\right\} = 1 + \mu_1t + \frac{\mu_2t^2}{2!} + \ldots + \frac{\mu_rt^r}{r!}. \quad (1.2.12.5b)$$

When it is substituted for t, (1.2.12.5b) is the characteristic function, or Fourier transform of $P(t)$ (Kendall & Stuart, 1958).

The first two terms in the exponent of (1.2.12.5a) can be omitted if the expansion is around the equilibrium position and the harmonic term is properly described by $P_0(\mathbf{u})$.

The Fourier transform of (1.2.12.5a) is, by analogy with the left-hand part of (1.2.12.5b) (with t replaced by $2\pi ih$),

$$T(\mathbf{H}) = \exp\left[\frac{(2\pi i)^3}{3!}\kappa^{jkl}h_jh_kh_l + \frac{(2\pi i)^4}{4!}\kappa^{jklm}h_jh_kh_lh_m + \ldots\right]T_0(\mathbf{H}) = \exp\left[-\frac{4}{3}\pi^3i\kappa^{jkl}h_jh_kh_l + \frac{2}{3}\pi^4\kappa^{jklm}h_jh_kh_lh_m + \ldots\right]T_0(\mathbf{H}), \quad (1.2.12.6)$$

where the first two terms have been omitted. Expression (1.2.12.6) is similar to (1.2.12.4) except that the entire series is in the exponent. Following Schwarzenbach (1986), (1.2.12.6) can be developed in a Taylor series, which gives

$$T(\mathbf{H}) = \left\{1 + \frac{(2\pi i)^3}{3!}\kappa^{jkl}h_jh_kh_l + \frac{(2\pi i)^4}{4!}\kappa^{jklm}h_jh_kh_lh_m + \ldots + \frac{(2\pi i)^6}{6!}\left[\kappa^{jklmp} + \frac{6!}{2!(3!)^2}\kappa^{jkl}\kappa^{mnp}\right]h_jh_kh_lh_mh_nh_p + \text{higher-order terms}\right\}T_0(\mathbf{H}). \quad (1.2.12.7)$$

This formulation, which is sometimes called the Edgeworth approximation (Zucker & Schulz, 1982), clearly shows the relation to the Gram–Charlier expansion (1.2.12.4), and corresponds to the probability distribution [analogous to (1.2.12.3)]

$$P(\mathbf{u}) = P_0(\mathbf{u})\left\{1 + \frac{1}{3!}\kappa^{jkl}H_{jkl}(\mathbf{u}) + \frac{1}{4!}\kappa^{jklm}H_{jklm}(\mathbf{u}) + \ldots + \frac{1}{6!}\left[\kappa^{jklmnp} + 10\kappa^{jkl}\kappa^{mnp}\right]H_{jklmnp} + \text{higher-order terms}\right\}. \quad (1.2.12.8)$$

The relation between the cumulants κ^{jkl} and the quasimoments c^{jkl} are apparent from comparison of (1.2.12.8) and (1.2.12.4):

$$\begin{aligned}
c^{jkl} &= \kappa^{jkl}\\
c^{jklm} &= \kappa^{jklm}\\
c^{jklmn} &= \kappa^{jklmn}\\
c^{jklmnp} &= \kappa^{jklmnp} + 10\kappa^{jkl}\kappa^{mnp}.
\end{aligned} \quad (1.2.12.9)$$

The sixth- and higher-order cumulants and quasimoments differ. Thus the third-order cumulant κ^{jkl} contributes not only to the coefficient of H_{jkl}, but also to higher-order terms of the probability distribution function. This is also the case for cumulants of higher orders. It implies that for a finite truncation of (1.2.12.6), the probability distribution cannot be represented by a finite number of terms. This is a serious difficulty when a probability distribution is to be derived from an experimental temperature factor of the cumulant type.

1.2.12.3. The one-particle potential (OPP) model

When an atom is considered as an independent oscillator vibrating in a potential well $V(\mathbf{u})$, its distribution may be described by Boltzmann statistics.

$$P(\mathbf{u}) = N \exp\{-V(\mathbf{u})/kT\}, \tag{1.2.12.10}$$

with N, the normalization constant, defined by $\int P(\mathbf{u})\ \mathrm{d}\mathbf{u} = 1$. The classical expression (1.2.12.10) is valid in the high-temperature limit for which $kT \gg V(\mathbf{u})$.

Following Dawson (1967) and Willis (1969), the potential function may be expanded in terms of increasing order of products of the contravariant displacement coordinates:

$$V = V_0 + \alpha_j u^j + \beta_{jk} u^j u^k + \gamma_{jkl} u^j u^k u^l + \delta_{jklm} u^j u^k u^l u^m + \ldots. \tag{1.2.12.11}$$

The equilibrium condition gives $\alpha_j = 0$. Substitution into (1.2.12.10) leads to an expression which may be simplified by the assumption that the leading term is the harmonic component represented by β_{jk}:

$$P(\mathbf{u}) = N \exp\{-\beta'_{jk} u^j u^k\} \times \{1 - \gamma'_{jkl} u^j u^k u^l - \delta'_{jklm} u^j u^k u^l u^m - \ldots\}, \tag{1.2.12.12}$$

in which $\beta' = \beta/kT$ *etc.* and the normalization factor N depends on the level of truncation.

The probability distribution is related to the spherical harmonic expansion. The ten products of the displacement parameters $u^j u^k u^l$, for example, are linear combinations of the seven octapoles ($l = 3$) and three dipoles ($l = 1$) (Coppens, 1980). The thermal probability distribution and the aspherical atom description can be separated only because the latter is essentially confined to the valence shell, while the former applies to all electrons which follow the nuclear motion in the atomic scattering model.

The Fourier transform of the OPP distribution, in a general coordinate system, is (Johnson, 1970*a*; Scheringer, 1985*a*)

$$T(\mathbf{H}) = T_0(\mathbf{H})\left[1 - \frac{4}{3}\pi^3 i\gamma'_{jkl} G^{jkl}(\mathbf{H}) + \frac{2}{3}\pi^4 \delta'_{jklm} G^{jklm}(\mathbf{H}) + \frac{4}{15}\pi^5 i\varepsilon'_{jklmn} G^{jklmn}(\mathbf{H}) - \frac{4}{45}\pi^6 i\varphi'_{jklmnp} G^{jklmnp}(\mathbf{H}) \ldots\right], \tag{1.2.12.13}$$

where T_0 is the harmonic temperature factor and G represents the Hermite polynomials in reciprocal space.

If the OPP temperature factor is expanded in the coordinate system which diagonalizes β_{jk}, simpler expressions are obtained in which the Hermite polynomials are replaced by products of the displacement coordinates u^j (Dawson *et al.*, 1967; Coppens, 1980; Tanaka & Marumo, 1983).

1.2.12.4. Relative merits of the three expansions

The relative merits of the Gram–Charlier and Edgeworth expansions have been discussed by Zucker & Schulz (1982), Kuhs (1983), and by Scheringer (1985*b*). In general, the Gram–Charlier expression is found to be preferable because it gives a better fit in the cases tested, and because its truncation is equivalent in real and reciprocal space. The latter is also true for the one-particle potential model, which is mathematically related to the Gram–Charlier expansion by the interchange of the real- and reciprocal-space expressions. The terms of the OPP model have a specific physical meaning. The model allows prediction of the temperature dependence of the temperature factor (Willis, 1969; Coppens, 1980), provided the potential function itself can be assumed to be temperature independent.

It has recently been shown that the Edgeworth expansion (1.2.12.5*a*) always has negative regions (Scheringer, 1985*b*). This implies that it is not a realistic description of a vibrating atom.

1.2.13. The generalized structure factor

In the generalized structure-factor formalism developed by Dawson (1975), the complex nature of both the atomic scattering factor and the generalized temperature factor are taken into account. We write for the atomic scattering factor:

$$f_j(\mathbf{H}) = f_{j,\mathrm{c}}(\mathbf{H}) + if_{j,\mathrm{a}}(\mathbf{H}) + f'_j + if''_j \tag{1.2.13.1a}$$

$$T_j(\mathbf{H}) = T_{j,\mathrm{c}}(\mathbf{H}) + iT_{j,\mathrm{a}}(\mathbf{H}) \tag{1.2.13.1b}$$

and

$$F(\mathbf{H}) = A(\mathbf{H}) + iB(\mathbf{H}), \tag{1.2.13.2}$$

where the subscripts c and a refer to the centrosymmetric and noncentrosymmetric components of the underlying electron distribution, respectively. Substitution in (1.2.4.2) gives for the real and imaginary components A and B of $F(\mathbf{H})$

$$A(\mathbf{H}) = \sum_j (f_{j,\mathrm{c}} + f'_j)[\cos(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{c} - \sin(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{a}] - (f_{j,\mathrm{a}} + f''_j)[\cos(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{a} + \sin(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{c}] \tag{1.2.13.3a}$$

and

$$B(\mathbf{H}) = \sum_j (f_{j,\mathrm{c}} + f'_j)[\cos(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{a} + \sin(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{c}] + (f_{j,\mathrm{a}} + f''_j)[\cos(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{c} - \sin(2\pi\mathbf{H}\cdot\mathbf{r}_j)T_\mathrm{a}] \tag{1.2.13.3b}$$

(McIntyre *et al.*, 1980; Dawson, 1967).

Expressions (1.2.13.3) illustrate the relation between valence-density anisotropy and anisotropy of thermal motion.

1.2.14. Conclusion

This chapter summarizes mathematical developments of the structure-factor formalism. The introduction of atomic asphericity into the formalism and the treatment of thermal motion are interlinked. It is important that the complexities of the thermal probability distribution function can often be reduced by very low temperature experimentation. Results obtained with the multipole formalism for atomic asphericity can be used to derive physical properties and *d*-orbital populations of transition-metal atoms (*IT* C, 2004). In such applications, the deconvolution of the charge density and the thermal vibrations is essential. This deconvolution is dependent on the adequacy of the models summarized here.

The author would like to thank several of his colleagues who gave invaluable criticism of earlier versions of this manuscript. Corrections and additions were made following comments by P. J. Becker, D. Feil, N. K. Hansen, G. McIntyre, E. N. Maslen, S. Ohba, C. Scheringer and D. Schwarzenbach. Z. Su contributed to the revised version of the manuscript. Support of this work by the US

National Science Foundation (CHE8711736 and CHE9317770) is gratefully acknowledged.

References

Arfken, G. (1970). *Mathematical Models for Physicists*, 2nd ed. New York, London: Academic Press.

Avery, J. & Ørmen, P.-J. (1979). *Generalized scattering factors and generalized Fourier transforms. Acta Cryst.* A**35**, 849–851.

Avery, J. & Watson, K. J. (1977). *Generalized X-ray scattering factors. Simple closed-form expressions for the one-centre case with Slater-type orbitals. Acta Cryst.* A**33**, 679–680.

Bentley, J. & Stewart, R. F. (1973). *Two-centre calculations for X-ray scattering. J. Comput. Phys.* **11**, 127–145.

Born, M. (1926). *Quantenmechanik der Stoszvorgänge. Z. Phys.* **38**, 803.

Clementi, E. & Raimondi, D. L. (1963). *Atomic screening constants from SCF functions. J. Chem. Phys.* **38**, 2686–2689.

Clementi, E. & Roetti, C. (1974). *Roothaan–Hartree–Fock atomic wavefunctions. At. Data Nucl. Data Tables*, **14**, 177–478.

Cohen-Tannoudji, C., Diu, B. & Laloe, F. (1977). *Quantum Mechanics.* New York: John Wiley and Paris: Hermann.

Condon, E. V. & Shortley, G. H. (1957). *The theory of atomic spectra.* London, New York: Cambridge University Press.

Coppens, P. (1980). *Thermal smearing and chemical bonding*. In *Electron and Magnetization Densities in Molecules and Solids*, edited by P. J. Becker, pp. 521–544. New York: Plenum.

Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Net atomic charges and molecular dipole moments from spherical-atom X-ray refinements, and the relation between atomic charges and shape. Acta Cryst.* A**35**, 63–72.

Coulson, C. A. (1961). *Valence*. Oxford University Press.

Cruickshank, D. W. J. (1956). *The analysis of the anisotropic thermal motion of molecules in crystals. Acta Cryst.* **9**, 754–756.

Dawson, B. (1967). *A general structure factor formalism for interpreting accurate X-ray and neutron diffraction data. Proc. R. Soc. London Ser. A*, **248**, 235–288.

Dawson, B. (1975). *Studies of atomic charge density by X-ray and neutron diffraction – a perspective*. In *Advances in Structure Research by Diffraction Methods.* Vol. 6, edited by W. Hoppe & R. Mason. Oxford: Pergamon Press.

Dawson, B., Hurley, A. C. & Maslen, V. W. (1967). *Anharmonic vibration in fluorite-structures. Proc. R. Soc. London Ser. A*, **298**, 289–306.

Dunitz, J. D. (1979). *X-ray analysis and the structure of organic molecules.* Ithaca, London: Cornell University Press.

Feil, D. (1977). *Diffraction physics. Isr. J. Chem.* **16**, 103–110.

Hansen, N. K. & Coppens, P. (1978). *Testing aspherical atom refinements on small-molecule data sets. Acta Cryst.* A**34**, 909–921.

Hehre, W. J., Ditchfield, R., Stewart, R. F. & Pople, J. A. (1970). *Self-consistent molecular orbital methods. IV. Use of Gaussian expansions of Slater-type orbitals. Extension to second-row molecules. J. Chem. Phys.* **52**, 2769–2773.

Hehre, W. J., Stewart, R. F. & Pople, J. A. (1969). *Self-consistent molecular orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. J. Chem. Phys.* **51**, 2657–2664.

Hirshfeld, F. L. (1977). *A deformation density refinement program. Isr. J. Chem.* **16**, 226–229.

International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press.

James, R. W. (1982). *The optical principles of the diffraction of X-rays*. Woodbridge: Oxbow Press.

Johnson, C. K. (1969). *Addition of higher cumulants to the crystallographic structure-factor equation: a generalized treatment for thermal-motion effects. Acta Cryst.* A**25**, 187–194.

Johnson, C. K. (1970*a*). *Series expansion models for thermal motion*. ACA Program and Abstracts, 1970 Winter Meeting, Tulane University, p. 60.

Johnson, C. K. (1970*b*). *An introduction to thermal-motion analysis*. In *Crystallographic Computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 207–219. Copenhagen: Munksgaard.

Johnson, C. K. & Levy, H. A. (1974). *Thermal motion analysis using Bragg diffraction data*. In *International Tables for X-ray Crystallography* (1974), Vol. IV, pp. 311–336. Birmingham: Kynoch Press.

Kara, M. & Kurki-Suonio, K. (1981). *Symmetrized multipole analysis of orientational distributions. Acta Cryst.* A**37**, 201–210.

Kendall, M. G. & Stuart, A. (1958). *The Advanced Theory of Statistics*. London: Griffin.

Kuhs, W. F. (1983). *Statistical description of multimodal atomic probability structures. Acta Cryst.* A**39**, 148–158.

Kurki-Suonio, K. (1977). *Symmetry and its implications. Isr. J. Chem.* **16**, 115–123.

Kutznetsov, P. I., Stratonovich, R. L. & Tikhonov, V. I. (1960). *Theory Probab. Its Appl. (USSR)*, **5**, 80–97.

Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures*. London: Bell.

McIntyre, G. J., Moss, G. & Barnea, Z. (1980). *Anharmonic temperature factors of zinc selenide determined by X-ray diffraction from an extended-face crystal. Acta Cryst.* A**36**, 482–490.

Materlik, G., Sparks, C. J. & Fischer, K. (1994). *Resonant Anomalous X-ray Scattering. Theory and Applications.* Amsterdam: North-Holland.

Paturle, A. & Coppens, P. (1988). *Normalization factors for spherical harmonic density functions. Acta Cryst.* A**44**, 6–7.

Ruedenberg, K. (1962). *The nature of the chemical bond. Phys. Rev.* **34**, 326–376.

Scheringer, C. (1985*a*). *A general expression for the anharmonic temperature factor in the isolated-atom-potential approach. Acta Cryst.* A**41**, 73–79.

Scheringer, C. (1985*b*). *A deficiency of the cumulant expansion of the anharmonic temperature factor. Acta Cryst.* A**41**, 79–81.

Schomaker, V. & Trueblood, K. N. (1968). *On the rigid-body motion of molecules in crystals. Acta Cryst.* B**24**, 63–76.

Schwarzenbach, D. (1986). Private communication.

Squires, G. L. (1978). *Introduction to the Theory of Thermal Neutron Scattering*. Cambridge University Press.

Stewart, R. F. (1969*a*). *Generalized X-ray scattering factors. J. Chem. Phys.* **51**, 4569–4577.

Stewart, R. F. (1969*b*). *Small Gaussian expansions of atomic orbitals. J. Chem. Phys.* **50**, 2485–2495.

Stewart, R. F. (1970). *Small Gaussian expansions of Slater-type orbitals. J. Chem. Phys.* **52**, 431–438.

Stewart, R. F. (1980). *Electron and Magnetization Densities in Molecules and Solids*, edited by P. J. Becker, pp. 439–442. New York: Plenum.

Stewart, R. F. & Hehre, W. J. (1970). *Small Gaussian expansions of atomic orbitals: second-row atoms. J. Chem. Phys.* **52**, 5243–5247.

Su, Z. & Coppens, P. (1990). *Closed-form expressions for Fourier–Bessel transforms of Slater-type functions. J. Appl. Cryst.* **23**, 71–73.

Su, Z. & Coppens, P. (1994). *Normalization factors for Kubic harmonic density functions. Acta Cryst.* A**50**, 408–409.

Tanaka, K. & Marumo, F. (1983). *Willis formalism of anharmonic temperature factors for a general potential and its application in the least-squares method. Acta Cryst.* A**39**, 631–641.

Von der Lage, F. C. & Bethe, H. A. (1947). *A method for obtaining electronic functions and eigenvalues in solids with an application to sodium. Phys. Rev.* **71**, 612–622.

Waller, I. & Hartree, D. R. (1929). *Intensity of total scattering X-rays. Proc. R. Soc. London Ser. A*, **124**, 119–142.

Weiss, R. J. & Freeman, A. J. (1959). *X-ray and neutron scattering for electrons in a crystalline field and the determination of outer electron configurations in iron and nickel. J. Phys. Chem. Solids*, **10**, 147–161.

Willis, B. T. M. (1969). *Lattice vibrations and the accurate determination of structure factors for the elastic scattering of X-rays and neutrons. Acta Cryst.* A**25**, 277–300.

Zucker, U. H. & Schulz, H. (1982). *Statistical approaches for the treatment of anharmonic motion in crystals. I. A comparison of the most frequently used formalisms of anharmonic thermal vibrations. Acta Cryst.* A**38**, 563–568.

International Tables for Crystallography (2010). Vol. B, Chapter 1.3, pp. 24–113.

1.3. Fourier transforms in crystallography: theory, algorithms and applications

By G. Bricogne

1.3.1. General introduction

Since the publication of Volume II of *International Tables*, most aspects of the theory, computation and applications of Fourier transforms have undergone considerable development, often to the point of being hardly recognizable.

The mathematical analysis of the Fourier transformation has been extensively reformulated within the framework of distribution theory, following Schwartz's work in the early 1950s.

The computation of Fourier transforms has been revolutionized by the advent of digital computers and of the Cooley–Tukey algorithm, and progress has been made at an ever-accelerating pace in the design of new types of algorithms and in optimizing their interplay with machine architecture.

These advances have transformed both theory and practice in several fields which rely heavily on Fourier methods; much of electrical engineering, for instance, has become digital signal processing.

By contrast, crystallography has remained relatively unaffected by these developments. From the conceptual point of view, old-fashioned Fourier series are still adequate for the quantitative description of X-ray diffraction, as this rarely entails consideration of molecular transforms between reciprocal-lattice points. From the practical point of view, three-dimensional Fourier transforms have mostly been used as a tool for visualizing electron-density maps, so that only moderate urgency was given to trying to achieve ultimate efficiency in these relatively infrequent calculations.

Recent advances in phasing and refinement methods, however, have placed renewed emphasis on concepts and techniques long used in digital signal processing, *e.g.* flexible sampling, Shannon interpolation, linear filtering, and interchange between convolution and multiplication. These methods are iterative in nature, and thus generate a strong incentive to design new crystallographic Fourier transform algorithms making the fullest possible use of all available symmetry to save both storage and computation.

As a result, need has arisen for a modern and coherent account of Fourier transform methods in crystallography which would provide:

(i) a simple and foolproof means of switching between the three different guises in which the Fourier transformation is encountered (Fourier transforms, Fourier series and discrete Fourier transforms), both formally and computationally;

(ii) an up-to-date presentation of the most important algorithms for the efficient numerical calculation of discrete Fourier transforms;

(iii) a systematic study of the incorporation of symmetry into the calculation of crystallographic discrete Fourier transforms;

(iv) a survey of the main types of crystallographic computations based on the Fourier transformation.

The rapid pace of progress in these fields implies that such an account would be struck by quasi-immediate obsolescence if it were written solely for the purpose of compiling a catalogue of results and formulae 'customized' for crystallographic use. Instead, the emphasis has been placed on a mode of presentation in which most results and formulae are *derived* rather than listed. This does entail a substantial mathematical overhead, but has the advantage of preserving in its 'native' form the context within which these results are obtained. It is this context, rather than any particular set of results, which constitutes the most fertile source of new ideas and new applications, and as such can have any hope at all of remaining useful in the long run.

These conditions have led to the following choices:

(i) the mathematical theory of the Fourier transformation has been cast in the language of Schwartz's theory of distributions which has long been adopted in several applied fields, in particular electrical engineering, with considerable success; the extra work involved handsomely pays for itself by allowing the three different types of Fourier transformations to be treated together, and by making all properties of the Fourier transform consequences of a single property (the convolution theorem). This is particularly useful in all questions related to the sampling theorem;

(ii) the various numerical algorithms have been presented as the consequences of basic algebraic phenomena involving Abelian groups, rings and finite fields; this degree of formalization greatly helps the subsequent incorporation of symmetry;

(iii) the algebraic nature of space groups has been re-emphasized so as to build up a framework which can accommodate both the phenomena used to factor the discrete Fourier transform and those which underlie the existence (and lead to the classification) of space groups; this common ground is found in the notion of *module over a group ring* (*i.e.* integral representation theory), which is then applied to the formulation of a large number of algorithms, many of which are new;

(iv) the survey of the main types of crystallographic computations has tried to highlight the roles played by various properties of the Fourier transformation, and the ways in which a better exploitation of these properties has been the driving force behind the discovery of more powerful methods.

In keeping with this philosophy, the theory is presented first, followed by the crystallographic applications. There are 'forward references' from mathematical results to the applications which later invoke them (thus giving 'real-life' examples rather than artificial ones), and 'backward references' as usual. In this way, the internal logic of the mathematical developments – the surest guide to future innovations – can be preserved, whereas the alternative solution of relegating these to appendices tends on the contrary to obscure that logic by subordinating it to that of the applications.

It is hoped that this attempt at an overall presentation of the main features of Fourier transforms and of their ubiquitous role in crystallography will be found useful by scientists both within and outside the field.

1.3.2. The mathematical theory of the Fourier transformation

1.3.2.1. Introduction

The Fourier transformation and the practical applications to which it gives rise occur in three different forms which, although they display a similar range of phenomena, normally require distinct formulations and different proof techniques:

(i) *Fourier transforms*, in which both function and transform depend on continuous variables;

(ii) *Fourier series*, which relate a periodic function to a discrete set of coefficients indexed by n-tuples of integers;

(iii) *discrete Fourier transforms*, which relate finite-dimensional vectors by linear operations representable by matrices.

Copyright © 2010 International Union of Crystallography

At the same time, the most useful property of the Fourier transformation – the exchange between multiplication and convolution – is mathematically the most elusive and the one which requires the greatest caution in order to avoid writing down meaningless expressions.

It is the unique merit of Schwartz's theory of distributions (Schwartz, 1966) that it affords complete control over all the troublesome phenomena which had previously forced mathematicians to settle for a piecemeal, fragmented theory of the Fourier transformation. By its ability to handle rigorously highly 'singular' objects (especially δ-functions, their derivatives, their tensor products, their products with smooth functions, their translates and lattices of these translates), distribution theory can deal with all the major properties of the Fourier transformation as particular instances of a single basic result (the exchange between multiplication and convolution), *and* can at the same time accommodate the three previously distinct types of Fourier theories within a unique framework. This brings great simplification to matters of central importance in crystallography, such as the relations between

(*a*) periodization, and sampling or decimation;

(*b*) Shannon interpolation, and masking by an indicator function;

(*c*) section, and projection;

(*d*) differentiation, and multiplication by a monomial;

(*e*) translation, and phase shift.

All these properties become subsumed under the same theorem.

This striking synthesis comes at a slight price, which is the relative complexity of the notion of distribution. It is first necessary to establish the notion of topological vector space and to gain sufficient control (or, at least, understanding) over convergence behaviour in certain of these spaces. The key notion of *metrizability* cannot be circumvented, as it underlies most of the constructs and many of the proof techniques used in distribution theory. Most of Section 1.3.2.2 builds up to the fundamental result at the end of Section 1.3.2.2.6.2, which is basic to the definition of a distribution in Section 1.3.2.3.4 and to all subsequent developments.

The reader mostly interested in applications will probably want to reach this section by starting with his or her favourite topic in Section 1.3.4, and following the backward references to the relevant properties of the Fourier transformation, then to the proof of these properties, and finally to the definitions of the objects involved. Hopefully, he or she will then feel inclined to follow the forward references and thus explore the subject from the abstract to the practical. The books by Dieudonné (1969) and Lang (1965) are particularly recommended as general references for all aspects of analysis and algebra.

1.3.2.2. Preliminary notions and notation

Throughout this text, $\mathbb{R}$ will denote the set of real numbers, $\mathbb{Z}$ the set of rational (signed) integers and $\mathbb{N}$ the set of natural (unsigned) integers. The symbol $\mathbb{R}^n$ will denote the Cartesian product of n copies of $\mathbb{R}$:

$$\mathbb{R}^n = \mathbb{R} \times \ldots \times \mathbb{R} \quad (n \text{ times}, n \geq 1),$$

so that an element $\mathbf{x}$ of $\mathbb{R}^n$ is an n-tuple of real numbers:

$$\mathbf{x} = (x_1, \ldots, x_n).$$

Similar meanings will be attached to $\mathbb{Z}^n$ and $\mathbb{N}^n$.

The symbol $\mathbb{C}$ will denote the set of complex numbers. If $z \in \mathbb{C}$, its modulus will be denoted by $|z|$, its conjugate by $\bar{z}$ (not z^*), and its real and imaginary parts by $\mathscr{Re}(z)$ and $\mathscr{Im}(z)$:

$$\mathscr{Re}(z) = \tfrac{1}{2}(z + \bar{z}), \qquad \mathscr{Im}(z) = \frac{1}{2i}(z - \bar{z}).$$

If X is a finite set, then $|X|$ will denote the number of its elements. If mapping f sends an element x of set X to the element $f(x)$ of set Y, the notation

$$f : x \longmapsto f(x)$$

will be used; the plain arrow $\rightarrow$ will be reserved for denoting limits, as in

$$\lim_{\rho \to \infty} \left(1 + \frac{x}{p}\right)^p = e^x.$$

If X is any set and S is a subset of X, the *indicator function* χ_s of S is the real-valued function on X defined by

$$\begin{aligned} \chi_S(x) &= 1 \quad \text{if } x \in S \\ &= 0 \quad \text{if } x \notin S. \end{aligned}$$

1.3.2.2.1. Metric and topological notions in $\mathbb{R}^n$

The set $\mathbb{R}^n$ can be endowed with the structure of a vector space of dimension n over $\mathbb{R}$, and can be made into a Euclidean space by treating its standard basis as an orthonormal basis and defining the Euclidean norm:

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

By misuse of notation, $\mathbf{x}$ will sometimes also designate the column vector of coordinates of $\mathbf{x} \in \mathbb{R}^n$; if these coordinates are referred to an orthonormal basis of $\mathbb{R}^n$, then

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2},$$

where $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$.

The distance between two points $\mathbf{x}$ and $\mathbf{y}$ defined by $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ allows the topological structure of $\mathbb{R}$ to be transferred to $\mathbb{R}^n$, making it a *metric space*. The basic notions in a metric space are those of neighbourhoods, of open and closed sets, of limit, of continuity, and of convergence (see Section 1.3.2.2.6.1).

A subset S of $\mathbb{R}^n$ is *bounded* if $\sup \|\mathbf{x} - \mathbf{y}\| < \infty$ as $\mathbf{x}$ and $\mathbf{y}$ run through S; it is *closed* if it contains the limits of all convergent sequences with elements in S. A subset K of $\mathbb{R}^n$ which is both bounded and closed has the property of being *compact*, *i.e.* that whenever K has been covered by a family of open sets, a finite subfamily can be found which suffices to cover K. Compactness is a very useful topological property for the purpose of proof, since it allows one to reduce the task of examining infinitely many local situations to that of examining only finitely many of them.

1.3.2.2.2. Functions over $\mathbb{R}^n$

Let φ be a complex-valued function over $\mathbb{R}^n$. The *support* of φ, denoted Supp φ, is the smallest closed subset of $\mathbb{R}^n$ outside which φ vanishes identically. If Supp φ is compact, φ is said to have compact support.

If $\mathbf{t} \in \mathbb{R}^n$, the *translate* of φ by $\mathbf{t}$, denoted $\tau_{\mathbf{t}}\varphi$, is defined by

$$(\tau_{\mathbf{t}}\varphi)(\mathbf{x}) = \varphi(\mathbf{x} - \mathbf{t}).$$

Its support is the geometric translate of that of φ:

$$\text{Supp } \tau_{\mathbf{t}}\varphi = \{\mathbf{x} + \mathbf{t} | \mathbf{x} \in \text{Supp } \varphi\}.$$

If A is a nonsingular linear transformation in $\mathbb{R}^n$, the *image* of φ by A, denoted $A^{\#}\varphi$, is defined by

$$(A^{\#}\varphi)(\mathbf{x}) = \varphi[A^{-1}(\mathbf{x})].$$

Its support is the geometric image of Supp φ under A:

$$\text{Supp } A^{\#}\varphi = \{A(\mathbf{x}) | \mathbf{x} \in \text{Supp } \varphi\}.$$

If S is a nonsingular affine transformation in $\mathbb{R}^n$ of the form

$$S(\mathbf{x}) = A(\mathbf{x}) + \mathbf{b}$$

with A linear, the image of φ by S is $S^{\#}\varphi = \tau_{\mathbf{b}}(A^{\#}\varphi)$, *i.e.*

$$(S^{\#}\varphi)(\mathbf{x}) = \varphi[A^{-1}(\mathbf{x} - \mathbf{b})].$$

Its support is the geometric image of Supp φ under S:

$$\text{Supp } S^{\#}\varphi = \{S(\mathbf{x}) | \mathbf{x} \in \text{Supp } \varphi\}.$$

It may be helpful to visualize the process of forming the image of a function by a geometric operation as consisting of applying that operation to the *graph* of that function, which is equivalent to applying the *inverse* transformation to the coordinates $\mathbf{x}$. This use of the inverse later affords the 'left-representation property' [see Section 1.3.4.2.2.2(*e*)] when the geometric operations form a group, which is of fundamental importance in the treatment of crystallographic symmetry (Sections 1.3.4.2.2.4, 1.3.4.2.2.5).

1.3.2.2.3. Multi-index notation

When dealing with functions in n variables and their derivatives, considerable abbreviation of notation can be obtained through the use of multi-indices.

A *multi-index* $\mathbf{p} \in \mathbb{N}^n$ is an n-tuple of natural integers: $\mathbf{p} = (p_1, \ldots, p_n)$. The *length* of $\mathbf{p}$ is defined as

$$|\mathbf{p}| = \sum_{i=1}^{n} p_i,$$

and the following abbreviations will be used:

(i) $\mathbf{x}^{\mathbf{p}} = x_1^{p_1} \ldots x_n^{p_n}$

(ii) $D_i f = \dfrac{\partial f}{\partial x_i} = \partial_i f$

(iii) $D^{\mathbf{p}} f = D_1^{p_1} \ldots D_n^{p_n} f = \dfrac{\partial^{|\mathbf{p}|} f}{\partial x_1^{p_1} \ldots \partial x_n^{p_n}}$

(iv) $\mathbf{q} \leq \mathbf{p}$ if and only if $q_i \leq p_i$ for all $i = 1, \ldots, n$

(v) $\mathbf{p} - \mathbf{q} = (p_1 - q_1, \ldots, p_n - q_n)$

(vi) $\mathbf{p}! = p_1! \times \ldots \times p_n!$

(vii) $\dbinom{\mathbf{p}}{\mathbf{q}} = \dbinom{p_1}{q_1} \times \ldots \times \dbinom{p_n}{q_n}$.

Leibniz's formula for the repeated differentiation of products then assumes the concise form

$$D^{\mathbf{p}}(fg) = \sum_{\mathbf{q} \leq \mathbf{p}} \binom{\mathbf{p}}{\mathbf{q}} D^{\mathbf{p}-\mathbf{q}} f D^{\mathbf{q}} g,$$

while the Taylor expansion of f to order m about $\mathbf{x} = \mathbf{a}$ reads

$$f(\mathbf{x}) = \sum_{|\mathbf{p}| \leq m} \frac{1}{\mathbf{p}!} [D^{\mathbf{p}} f(\mathbf{a})](\mathbf{x} - \mathbf{a})^{\mathbf{p}} + o(\|\mathbf{x} - \mathbf{a}\|^m).$$

In certain sections the notation ∇f will be used for the gradient vector of f, and the notation $(\nabla\nabla^T)f$ for the Hessian matrix of its mixed second-order partial derivatives:

$$\nabla = \begin{pmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{pmatrix}, \quad \nabla f = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix},$$

$$(\nabla\nabla^T)f = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

1.3.2.2.4. Integration, L^p spaces

The Riemann integral used in elementary calculus suffers from the drawback that vector spaces of Riemann-integrable functions over $\mathbb{R}^n$ are not *complete* for the topology of convergence in the mean: a Cauchy sequence of integrable functions may converge to a nonintegrable function.

To obtain the property of completeness, which is fundamental in functional analysis, it was necessary to extend the notion of integral. This was accomplished by Lebesgue [see Berberian (1962), Dieudonné (1970), or Chapter 1 of Dym & McKean (1972) and the references therein, or Chapter 9 of Sprecher (1970)], and entailed identifying functions which differed only on a subset of zero measure in $\mathbb{R}^n$ (such functions are said to be equal 'almost everywhere'). The vector spaces $L^p(\mathbb{R}^n)$ consisting of function classes f modulo this identification for which

$$\|\mathbf{f}\|_p = \left(\int_{\mathbb{R}^n} |f(\mathbf{x})|^p \, \mathrm{d}^n\mathbf{x} \right)^{1/p} < \infty$$

are then complete for the topology induced by the norm $\|.\|_p$: the limit of every Cauchy sequence of functions in L^p is itself a function in L^p (Riesz–Fischer theorem).

The space $L^1(\mathbb{R}^n)$ consists of those function classes f such that

$$\|f\|_1 = \int_{\mathbb{R}^n} |f(\mathbf{x})| \, \mathrm{d}^n\mathbf{x} < \infty$$

which are called *summable* or *absolutely integrable*. The convolution product:

$$\begin{aligned}(f * g)(\mathbf{x}) &= \int_{\mathbb{R}^n} f(\mathbf{y}) g(\mathbf{x}-\mathbf{y}) \, \mathrm{d}^n\mathbf{y} \\ &= \int_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y}) g(\mathbf{y}) \, \mathrm{d}^n\mathbf{y} = (g * f)(\mathbf{x})\end{aligned}$$

is well defined; combined with the vector space structure of L^1, it makes L^1 into a (commutative) *convolution algebra*. However, this algebra has no unit element: there is no $f \in L^1$ such that $f * g = g$ for all $g \in L^1$; it has only approximate units, *i.e.* sequences (f_ν) such that $f_\nu * g$ tends to g in the L^1 topology as $\nu \to \infty$. This is one of the starting points of distribution theory.

The space $L^2(\mathbb{R}^n)$ of *square-integrable* functions can be endowed with a scalar product

$$(f, g) = \int_{\mathbb{R}^n} \overline{f(\mathbf{x})} g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x}$$

which makes it into a *Hilbert space*. The Cauchy–Schwarz inequality

$$|(f, g)| \leq [(f, f)(g, g)]^{1/2}$$

generalizes the fact that the absolute value of the cosine of an angle is less than or equal to 1.

The space $L^\infty(\mathbb{R}^n)$ is defined as the space of functions f such that

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p = \lim_{p\to\infty} \left(\int_{\mathbb{R}^n} |f(\mathbf{x})|^p \, \mathrm{d}^n\mathbf{x} \right)^{1/p} < \infty.$$

The quantity $\|f\|_\infty$ is called the 'essential sup norm' of f, as it is the smallest positive number which $|f(\mathbf{x})|$ exceeds only on a subset of zero measure in $\mathbb{R}^n$. A function $f \in L^\infty$ is called *essentially bounded*.

1.3.2.2.5. Tensor products. Fubini's theorem

Let $f \in L^1(\mathbb{R}^m)$, $g \in L^1(\mathbb{R}^n)$. Then the function

$$f \otimes g : (\mathbf{x}, \mathbf{y}) \longmapsto f(\mathbf{x}) g(\mathbf{y})$$

is called the *tensor product* of f and g, and belongs to $L^1(\mathbb{R}^m \times \mathbb{R}^n)$. The finite linear combinations of functions of the form $f \otimes g$ span a subspace of $L^1(\mathbb{R}^m \times \mathbb{R}^n)$ called the tensor product of $L^1(\mathbb{R}^m)$ and $L^1(\mathbb{R}^n)$ and denoted $L^1(\mathbb{R}^m) \otimes L^1(\mathbb{R}^n)$.

The integration of a general function over $\mathbb{R}^m \times \mathbb{R}^n$ may be accomplished in two steps according to *Fubini's theorem*. Given $F \in L^1(\mathbb{R}^m \times \mathbb{R}^n)$, the functions

$$\begin{aligned}F_1 &: \mathbf{x} \longmapsto \int_{\mathbb{R}^n} F(\mathbf{x}, \mathbf{y}) \, \mathrm{d}^n\mathbf{y} \\ F_2 &: \mathbf{y} \longmapsto \int_{\mathbb{R}^m} F(\mathbf{x}, \mathbf{y}) \, \mathrm{d}^m\mathbf{x}\end{aligned}$$

exist for almost all $\mathbf{x} \in \mathbb{R}^m$ and almost all $\mathbf{y} \in \mathbb{R}^n$, respectively, are integrable, and

$$\int_{\mathbb{R}^m\times\mathbb{R}^n} F(\mathbf{x}, \mathbf{y}) \, \mathrm{d}^m\mathbf{x} \, \mathrm{d}^n\mathbf{y} = \int_{\mathbb{R}^m} F_1(\mathbf{x}) \, \mathrm{d}^m\mathbf{x} = \int_{\mathbb{R}^n} F_2(\mathbf{y}) \, \mathrm{d}^n\mathbf{y}.$$

Conversely, if any one of the integrals

(i) $$\int_{\mathbb{R}^m\times\mathbb{R}^n} |F(\mathbf{x}, \mathbf{y})| \, \mathrm{d}^m\mathbf{x} \, \mathrm{d}^n\mathbf{y}$$

(ii) $$\int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} |F(\mathbf{x}, \mathbf{y})| \, \mathrm{d}^n\mathbf{y} \right) \mathrm{d}^m\mathbf{x}$$

(iii) $$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} |F(\mathbf{x}, \mathbf{y})| \, \mathrm{d}^m\mathbf{x} \right) \mathrm{d}^n\mathbf{y}$$

is finite, then so are the other two, and the identity above holds. It is then (and only then) permissible to change the order of integrations.

Fubini's theorem is of fundamental importance in the study of tensor products and convolutions of distributions.

1.3.2.2.6. Topology in function spaces

Geometric intuition, which often makes 'obvious' the topological properties of the real line and of ordinary space, cannot be relied upon in the study of function spaces: the latter are infinite-dimensional, and several inequivalent notions of convergence may exist. A careful analysis of topological concepts and of their interrelationship is thus a necessary prerequisite to the study of these spaces. The reader may consult Dieudonné (1969, 1970), Friedman (1970), Trèves (1967) and Yosida (1965) for detailed expositions.

1.3.2.2.6.1. General topology

Most topological notions are first encountered in the setting of *metric spaces*. A metric space E is a set equipped with a *distance function* d from $E \times E$ to the non-negative reals which satisfies:

(i)	$d(x, y) = d(y, x)$	$\forall x, y \in E$	(symmetry);
(ii)	$d(x, y) = 0$	iff $x = y$	(separation);
(iii)	$d(x, z) \leq d(x, y) + d(y, z)$	$\forall x, y, z \in E$	(triangular inequality).

By means of d, the following notions can be defined: open balls, neighbourhoods; open and closed sets, interior and closure; convergence of sequences, continuity of mappings; Cauchy sequences and completeness; compactness; connectedness. They suffice for the investigation of a great number of questions in analysis and geometry (see *e.g.* Dieudonné, 1969).

Many of these notions turn out to depend only on the properties of the collection $\mathcal{O}(E)$ of open subsets of E: two distance functions leading to the same $\mathcal{O}(E)$ lead to identical topological properties. An axiomatic reformulation of topological notions is thus possible: a *topology* in E is a collection $\mathcal{O}(E)$ of subsets of E which satisfy suitable axioms and are deemed open irrespective

of the way they are obtained. From the practical standpoint, however, a topology which can be obtained from a distance function (called a *metrizable* topology) has the very useful property that *the notions of closure, limit and continuity may be defined by means of sequences*. For nonmetrizable topologies, these notions are much more difficult to handle, requiring the use of 'filters' instead of sequences.

In some spaces E, a topology may be most naturally defined by a family of *pseudo-distances* $(d_\alpha)_{\alpha\in A}$, where each d_α satisfies (i) and (iii) but not (ii). Such spaces are called *uniformizable*. If for every pair $(x, y) \in E \times E$ there exists $\alpha \in A$ such that $d_\alpha(x, y) \neq 0$, then the separation property can be recovered. If furthermore a *countable* subfamily of the d_α suffices to define the topology of E, the latter can be shown to be *metrizable*, so that limiting processes in E may be studied by means of sequences.

1.3.2.2.6.2. Topological vector spaces

The function spaces E of interest in Fourier analysis have an underlying vector space structure over the field $\mathbb{C}$ of complex numbers. A topology on E is said to be *compatible* with a vector space structure on E if vector addition [*i.e.* the map $(\mathbf{x}, \mathbf{y}) \longmapsto \mathbf{x} + \mathbf{y}$] and scalar multiplication [*i.e.* the map $(\lambda, \mathbf{x}) \longmapsto \lambda\mathbf{x}$] are both *continuous*; E is then called a *topological vector space*. Such a topology may be defined by specifying a 'fundamental system S of neighbourhoods of $\mathbf{0}$', which can then be translated by vector addition to construct neighbourhoods of other points $\mathbf{x} \neq \mathbf{0}$.

A *norm* ν on a vector space E is a non-negative real-valued function on $E \times E$ such that

(i′)	$\nu(\lambda\mathbf{x}) = \lvert\lambda\rvert\nu(\mathbf{x})$	for all $\lambda \in \mathbb{C}$ and $\mathbf{x} \in E$;
(ii′)	$\nu(\mathbf{x}) = 0$	if and only if $\mathbf{x} = \mathbf{0}$;
(iii′)	$\nu(\mathbf{x} + \mathbf{y}) \leq \nu(\mathbf{x}) + \nu(\mathbf{y})$	for all $\mathbf{x}, \mathbf{y} \in E$.

Subsets of E defined by conditions of the form $\nu(\mathbf{x}) \leq r$ with $r > 0$ form a fundamental system of neighbourhoods of $\mathbf{0}$. The corresponding topology makes E a *normed space*. This topology is *metrizable*, since it is equivalent to that derived from the translation-invariant distance $d(\mathbf{x}, \mathbf{y}) = \nu(\mathbf{x} - \mathbf{y})$. Normed spaces which are *complete*, *i.e.* in which all Cauchy sequences converge, are called *Banach spaces*; they constitute the natural setting for the study of differential calculus.

A *semi-norm* σ on a vector space E is a positive real-valued function on $E \times E$ which satisfies (i′) and (iii′) but not (ii′). Given a set Σ of semi-norms on E such that any pair $(\mathbf{x}, \mathbf{y})$ in $E \times E$ is separated by at least one $\sigma \in \Sigma$, let B be the set of those subsets $\Gamma_{\sigma,r}$ of E defined by a condition of the form $\sigma(\mathbf{x}) \leq r$ with $\sigma \in \Sigma$ and $r > 0$; and let S be the set of finite intersections of elements of B. Then there exists a unique topology on E for which S is a fundamental system of neighbourhoods of $\mathbf{0}$. This topology is *uniformizable* since it is equivalent to that derived from the family of translation-invariant pseudo-distances $(\mathbf{x}, \mathbf{y}) \longmapsto \sigma(\mathbf{x} - \mathbf{y})$. It is *metrizable* if and only if it can be constructed by the above procedure with Σ a *countable* set of semi-norms. If furthermore E is complete, E is called a *Fréchet space*.

If E is a topological vector space over $\mathbb{C}$, its *dual* E^* is the set of all linear mappings from E to $\mathbb{C}$ (which are also called *linear forms*, or *linear functionals*, over E). The subspace of E^* consisting of all linear forms which are *continuous* for the topology of E is called the *topological dual* of E and is denoted E'. If the topology on E is metrizable, then the continuity of a linear form $T \in E'$ at $f \in E$ can be ascertained by means of sequences, *i.e.* by checking that the sequence $[T(f_j)]$ of complex numbers converges to $T(f)$ in $\mathbb{C}$ whenever the sequence (f_j) converges to f in E.

1.3.2.3. Elements of the theory of distributions

1.3.2.3.1. Origins

At the end of the 19th century, Heaviside proposed under the name of 'operational calculus' a set of rules for solving a class of differential, partial differential and integral equations encountered in electrical engineering (today's 'signal processing'). These rules worked remarkably well but were devoid of mathematical justification (see Whittaker, 1928). In 1926, Dirac introduced his famous δ-function [see Dirac (1958), pp. 58–61], which was found to be related to Heaviside's constructs. Other singular objects, together with procedures to handle them, had already appeared in several branches of analysis [Cauchy's 'principal values'; Hadamard's 'finite parts' (Hadamard, 1932, 1952); Riesz's regularization methods for certain divergent integrals (Riesz, 1938, 1949)] as well as in the theories of Fourier series and integrals (see *e.g.* Bochner, 1932, 1959). Their very definition often verged on violating the rigorous rules governing limiting processes in analysis, so that subsequent recourse to limiting processes could lead to erroneous results; *ad hoc* precautions thus had to be observed to avoid mistakes in handling these objects.

In 1945–1950, Laurent Schwartz proposed his theory of distributions (see Schwartz, 1966), which provided a unified and definitive treatment of all these questions, with a striking combination of rigour and simplicity. Schwartz's treatment of Dirac's δ-function illustrates his approach in a most direct fashion. Dirac's original definition reads:

(i) $\delta(\mathbf{x}) = 0$ for $\mathbf{x} \neq \mathbf{0}$,

(ii) $\int_{\mathbb{R}^n} \delta(\mathbf{x})\, \mathrm{d}^n\mathbf{x} = 1$.

These two conditions are irreconcilable with Lebesgue's theory of integration: by (i), δ vanishes almost everywhere, so that its integral in (ii) must be 0, not 1.

A better definition consists in specifying that

(iii) $\int_{\mathbb{R}^n} \delta(\mathbf{x})\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x} = \varphi(\mathbf{0})$

for any function φ sufficiently well behaved near $\mathbf{x} = \mathbf{0}$. This is related to the problem of finding a unit for convolution (Section 1.3.2.2.4). As will now be seen, this definition is still unsatisfactory. Let the sequence (f_ν) in $L^1(\mathbb{R}^n)$ be an approximate convolution unit, *e.g.*

$$f_\nu(\mathbf{x}) = \left(\frac{\nu}{2\pi}\right)^{1/2} \exp(-\tfrac{1}{2}\nu^2\|\mathbf{x}\|^2).$$

Then for any well behaved function φ the integrals

$$\int_{\mathbb{R}^n} f_\nu(\mathbf{x})\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x}$$

exist, and the sequence of their numerical values tends to $\varphi(\mathbf{0})$. It is tempting to combine this with (iii) to conclude that δ is the limit of the sequence (f_ν) as $\nu \to \infty$. However,

$$\lim f_\nu(\mathbf{x}) = 0 \quad \text{as } \nu \to \infty$$

almost everywhere in $\mathbb{R}^n$ and the crux of the problem is that

$$\varphi(\mathbf{0}) = \lim_{\nu\to\infty} \int_{\mathbb{R}^n} f_\nu(\mathbf{x})\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x}$$
$$\neq \int_{\mathbb{R}^n} \Big[\lim_{\nu\to\infty} f_\nu(\mathbf{x})\Big]\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x} = 0$$

because the sequence (f_ν) does not satisfy the hypotheses of Lebesgue's dominated convergence theorem.

Schwartz's solution to this problem is deceptively simple: the regular behaviour one is trying to capture is an attribute not of the sequence of functions (f_ν), but of the sequence of continuous linear functionals

$$T_\nu : \varphi \longmapsto \int_{\mathbb{R}^n} f_\nu(\mathbf{x})\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x}$$

which has as a limit the continuous functional

$$T : \varphi \longmapsto \varphi(\mathbf{0}).$$

It is the latter functional which constitutes the proper definition of δ. The previous paradoxes arose because one insisted on writing down the simple linear operation T in terms of an integral.

The essence of Schwartz's theory of distributions is thus that, rather than try to define and handle 'generalized functions' *via* sequences such as (f_ν) [an approach adopted *e.g.* by Lighthill (1958) and Erdélyi (1962)], one should instead look at them as continuous linear functionals over spaces of well behaved functions.

There are many books on distribution theory and its applications. The reader may consult in particular Schwartz (1965, 1966), Gel'fand & Shilov (1964), Bremermann (1965), Trèves (1967), Challifour (1972), Friedlander (1982), and the relevant chapters of Hörmander (1963) and Yosida (1965). Schwartz (1965) is especially recommended as an introduction.

1.3.2.3.2. Rationale

The guiding principle which leads to requiring that the functions φ above (traditionally called 'test functions') should be well behaved is that correspondingly 'wilder' behaviour can then be accommodated in the limiting behaviour of the f_ν while still keeping the integrals $\int_{\mathbb{R}^n} f_\nu\varphi\,\mathrm{d}^n\mathbf{x}$ under control. Thus

(i) to minimize restrictions on the limiting behaviour of the f_ν *at infinity*, the φ's will be chosen to have *compact support*;

(ii) to minimize restrictions on the *local* behaviour of the f_ν, the φ's will be chosen *infinitely differentiable*.

To ensure further the *continuity* of functionals such as T_ν with respect to the test function φ as the f_ν go increasingly wild, very strong control will have to be exercised in the way in which a sequence (φ_j) of test functions will be said to converge towards a limiting φ: conditions will have to be imposed not only on the values of the *functions* φ_j, but also on those of all their *derivatives*. Hence, defining a strong enough topology on the space of test functions φ is an *essential* prerequisite to the development of a satisfactory theory of distributions.

1.3.2.3.3. Test-function spaces

With this rationale in mind, the following function spaces will be defined for any open subset Ω of $\mathbb{R}^n$ (which may be the whole of $\mathbb{R}^n$):

(*a*) $\mathscr{E}(\Omega)$ is the space of complex-valued functions over Ω which are indefinitely differentiable;

(*b*) $\mathscr{D}(\Omega)$ is the subspace of $\mathscr{E}(\Omega)$ consisting of functions with (unspecified) compact support contained in $\mathbb{R}^n$;

(*c*) $\mathscr{D}_K(\Omega)$ is the subspace of $\mathscr{D}(\Omega)$ consisting of functions whose (compact) support is contained within a fixed compact subset K of Ω.

When Ω is unambiguously defined by the context, we will simply write $\mathscr{E}$, $\mathscr{D}$, $\mathscr{D}_K$.

It sometimes suffices to require the existence of continuous derivatives only up to finite order m inclusive. The corresponding spaces are then denoted $\mathscr{E}^{(m)}$, $\mathscr{D}^{(m)}$, $\mathscr{D}_K^{(m)}$ with the convention that if $m = 0$, only continuity is required.

The *topologies* on these spaces constitute the most important ingredients of distribution theory, and will be outlined in some detail.

1.3.2.3.3.1. Topology on $\mathscr{E}(\Omega)$

It is defined by the family of semi-norms

$$\varphi \in \mathscr{E}(\Omega) \longmapsto \sigma_{\mathbf{p},K}(\varphi) = \sup_{\mathbf{x}\in K} |D^{\mathbf{p}}\varphi(\mathbf{x})|,$$

where $\mathbf{p}$ is a multi-index and K a compact subset of Ω. A fundamental system S of neighbourhoods of the origin in $\mathscr{E}(\Omega)$ is given by subsets of $\mathscr{E}(\Omega)$ of the form

$$V(m, \varepsilon, K) = \{\varphi \in \mathscr{E}(\Omega) | |\mathbf{p}| \leq m \Rightarrow \sigma_{\mathbf{p},K}(\varphi) < \varepsilon\}$$

for all natural integers m, positive real ε, and compact subset K of Ω. Since a *countable* family of compact subsets K suffices to cover Ω, and since restricted values of ε of the form $\varepsilon = 1/N$ lead to the same topology, S is equivalent to a *countable* system of neighbourhoods and hence $\mathscr{E}(\Omega)$ *is metrizable*.

Convergence in $\mathscr{E}$ may thus be defined by means of sequences. A sequence (φ_ν) in $\mathscr{E}$ will be said to converge to 0 if for any given $V(m, \varepsilon, K)$ there exists ν_0 such that $\varphi_\nu \in V(m, \varepsilon, K)$ whenever $\nu > \nu_0$; in other words, if the φ_ν and all their derivatives $D^{\mathbf{p}}\varphi_\nu$ converge to 0 uniformly on any given compact K in Ω.

1.3.2.3.3.2. Topology on $\mathscr{D}_k(\Omega)$

It is defined by the family of semi-norms

$$\varphi \in \mathscr{D}_K(\Omega) \longmapsto \sigma_{\mathbf{p}}(\varphi) = \sup_{\mathbf{x}\in K} |D^{\mathbf{p}}\varphi(\mathbf{x})|,$$

where K is now fixed. The fundamental system S of neighbourhoods of the origin in $\mathscr{D}_K$ is given by sets of the form

$$V(m, \varepsilon) = \{\varphi \in \mathscr{D}_K(\Omega) | |\mathbf{p}| \leq m \Rightarrow \sigma_{\mathbf{p}}(\varphi) < \varepsilon\}.$$

It is equivalent to the countable subsystem of the $V(m, 1/N)$, hence $\mathscr{D}_K(\Omega)$ *is metrizable*.

Convergence in $\mathscr{D}_K$ may thus be defined by means of sequences. A sequence (φ_ν) in $\mathscr{D}_K$ will be said to converge to 0 if for any given $V(m, \varepsilon)$ there exists ν_0 such that $\varphi_\nu \in V(m, \varepsilon)$ whenever $\nu > \nu_0$; in other words, if the φ_ν and all their derivatives $D^{\mathbf{p}}\varphi_\nu$ converge to 0 uniformly in K.

1.3.2.3.3.3. Topology on $\mathscr{D}(\Omega)$

It is defined by the fundamental system of neighbourhoods of the origin consisting of sets of the form

$$V((m), (\varepsilon))$$
$$= \left\{\varphi \in \mathscr{D}(\Omega) | |\mathbf{p}| \leq m_\nu \Rightarrow \sup_{\|\mathbf{x}\|\leq\nu} |D^{\mathbf{p}}\varphi(\mathbf{x})| < \varepsilon_\nu \text{ for all } \nu\right\},$$

where (m) is an increasing sequence (m_ν) of integers tending to $+\infty$ and (ε) is a decreasing sequence (ε_ν) of positive reals tending to 0, as $\nu \to \infty$.

This topology is *not metrizable*, because the sets of sequences (m) and (ε) are essentially uncountable. It can, however, be shown to be the *inductive limit* of the topology of the subspaces $\mathscr{D}_K$, in the following sense: V is a neighbourhood of the origin in $\mathscr{D}$ if and only if its intersection with $\mathscr{D}_K$ is a neighbourhood of the origin in $\mathscr{D}_K$ for any given compact K in Ω.

A sequence (φ_ν) in $\mathscr{D}$ will thus be said to converge to 0 in $\mathscr{D}$ if all the φ_ν belong to some $\mathscr{D}_K$ (with K a compact subset of Ω independent of ν) and if (φ_ν) converges to 0 in $\mathscr{D}_K$.

As a result, a complex-valued functional T on $\mathscr{D}$ will be said to be continuous for the topology of $\mathscr{D}$ if and only if, for any given compact K in Ω, its restriction to $\mathscr{D}_K$ is continuous for the topology of $\mathscr{D}_K$, *i.e.* maps convergent sequences in $\mathscr{D}_K$ to convergent sequences in $\mathbb{C}$.

This property of $\mathscr{D}$, *i.e.* having a nonmetrizable topology which is the inductive limit of metrizable topologies in its subspaces $\mathscr{D}_K$, conditions the whole structure of distribution theory and dictates that of many of its proofs.

1.3.2.3.3.4. Topologies on $\mathscr{E}^{(m)}$, $\mathscr{D}_k^{(m)}$, $\mathscr{D}^{(m)}$

These are defined similarly, but only involve conditions on derivatives up to order m.

1.3.2.3.4. Definition of distributions

A *distribution* T on Ω is a *linear form* over $\mathscr{D}(\Omega)$, *i.e.* a map

$$T : \varphi \longmapsto \langle T, \varphi \rangle$$

which associates linearly a complex number $\langle T, \varphi \rangle$ to any $\varphi \in \mathscr{D}(\Omega)$, and which is *continuous* for the topology of that space. In the terminology of Section 1.3.2.2.6.2, T is an element of $\mathscr{D}'(\Omega)$, the *topological dual* of $\mathscr{D}(\Omega)$.

Continuity over $\mathscr{D}$ is equivalent to continuity over $\mathscr{D}_K$ for all compact K contained in Ω, and hence to the condition that for any sequence (φ_ν) in $\mathscr{D}$ such that

(i) Supp φ_ν is contained in some compact K independent of ν,

(ii) the sequences $(|D^{\mathbf{p}}\varphi_\nu|)$ converge uniformly to 0 on K for all multi-indices $\mathbf{p}$;

then the sequence of complex numbers $\langle T, \varphi_\nu \rangle$ converges to 0 in $\mathbb{C}$.

If the continuity of a distribution T requires (ii) for $|\mathbf{p}| \leq m$ only, T may be defined over $\mathscr{D}^{(m)}$ and thus $T \in \mathscr{D}'^{(m)}$; T is said to be a *distribution of finite order m*. In particular, for $m = 0$, $\mathscr{D}^{(0)}$ is the space of continuous functions with compact support, and a distribution $T \in \mathscr{D}'^{(0)}$ is a (Radon) *measure* as used in the theory of integration. Thus measures are particular cases of distributions.

Generally speaking, the *larger* a space of test functions, the *smaller* its topological dual:

$$m < n \Rightarrow \mathscr{D}^{(m)} \supset \mathscr{D}^{(n)} \Rightarrow \mathscr{D}'^{(n)} \supset \mathscr{D}'^{(m)}.$$

This clearly results from the observation that if the φ's are allowed to be less regular, then less wildness can be accommodated in T if the continuity of the map $\varphi \longmapsto \langle T, \varphi \rangle$ with respect to φ is to be preserved.

1.3.2.3.5. First examples of distributions

(i) The linear map $\varphi \longmapsto \langle \delta, \varphi \rangle = \varphi(\mathbf{0})$ is a measure (*i.e.* a zeroth-order distribution) called Dirac's measure or (improperly) Dirac's 'δ-function'.

(ii) The linear map $\varphi \longmapsto \langle \delta_{(\mathbf{a})}, \varphi \rangle = \varphi(\mathbf{a})$ is called Dirac's measure at point $\mathbf{a} \in \mathbb{R}^n$.

(iii) The linear map $\varphi \longmapsto (-1)^{\mathbf{p}} D^{\mathbf{p}}\varphi(\mathbf{a})$ is a distribution of order $m = |\mathbf{p}| > 0$, and hence is not a measure.

(iv) The linear map $\varphi \longmapsto \sum_{\nu>0} \varphi^{(\nu)}(\nu)$ is a distribution of infinite order on $\mathbb{R}$: the order of differentiation is bounded for each φ (because φ has compact support) but is not as φ varies.

(v) If $(\mathbf{p}_\nu)$ is a sequence of multi-indices $\mathbf{p}_\nu = (p_{1\nu}, \ldots, p_{n\nu})$ such that $|\mathbf{p}_\nu| \to \infty$ as $\nu \to \infty$, then the linear map $\varphi \longmapsto \sum_{\nu>0} (D^{\mathbf{p}_\nu}\varphi)(\mathbf{p}_\nu)$ is a distribution of infinite order on $\mathbb{R}^n$.

1.3.2.3.6. Distributions associated to locally integrable functions

Let f be a complex-valued function over Ω such that $\int_K |f(\mathbf{x})|\, \mathrm{d}^n\mathbf{x}$ exists for any given compact K in Ω; f is then called *locally integrable*.

The linear mapping from $\mathscr{D}(\Omega)$ to $\mathbb{C}$ defined by

$$\varphi \longmapsto \int_\Omega f(\mathbf{x})\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x}$$

may then be shown to be continuous over $\mathscr{D}(\Omega)$. It thus defines a *distribution* $T_f \in \mathscr{D}'(\Omega)$:

$$\langle T_f, \varphi \rangle = \int_\Omega f(\mathbf{x})\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x}.$$

As the continuity of T_f only requires that $\varphi \in \mathscr{D}^{(0)}(\Omega)$, T_f is actually a Radon measure.

It can be shown that two locally integrable functions f and g define the same distribution, *i.e.*

$$\langle T_f, \varphi \rangle = \langle T_K, \varphi \rangle \quad \text{for all } \varphi \in \mathscr{D},$$

if and only if they are equal almost everywhere. The classes of locally integrable functions modulo this equivalence form a vector space denoted $L^1_{\mathrm{loc}}(\Omega)$; each element of $L^1_{\mathrm{loc}}(\Omega)$ may therefore be identified with the distribution T_f defined by any one of its representatives f.

1.3.2.3.7. Support of a distribution

A distribution $T \in \mathscr{D}'(\Omega)$ is said to *vanish* on an open subset ω of Ω if it vanishes on all functions in $\mathscr{D}(\omega)$, *i.e.* if $\langle T, \varphi \rangle = 0$ whenever $\varphi \in \mathscr{D}(\omega)$.

The *support* of a distribution T, denoted Supp T, is then defined as the complement of the set-theoretic union of those open subsets ω on which T vanishes; or equivalently as the smallest closed subset of Ω outside which T vanishes.

When $T = T_f$ for $f \in L^1_{\mathrm{loc}}(\Omega)$, then Supp T = Supp f, so that the two notions coincide. Clearly, if Supp T and Supp φ are disjoint subsets of Ω, then $\langle T, \varphi \rangle = 0$.

It can be shown that any distribution $T \in \mathscr{D}'$ with compact support may be extended from $\mathscr{D}$ to $\mathscr{E}$ while remaining continuous, so that $T \in \mathscr{E}'$; and that conversely, if $S \in \mathscr{E}'$, then its restriction T to $\mathscr{D}$ is a distribution with compact support. Thus, *the topological dual $\mathscr{E}'$ of $\mathscr{E}$ consists of those distributions in $\mathscr{D}'$ which have compact support*. This is intuitively clear since, if the condition of having compact support is fulfilled by T, it needs no longer be required of φ, which may then roam through $\mathscr{E}$ rather than $\mathscr{D}$.

1.3.2.3.8. Convergence of distributions

A *sequence* (T_j) of distributions will be said to converge in $\mathscr{D}'$ to a distribution T as $j \to \infty$ if, for any given $\varphi \in \mathscr{D}$, the sequence of complex numbers $(\langle T_j, \varphi \rangle)$ converges in $\mathbb{C}$ to the complex number $\langle T, \varphi \rangle$.

A *series* $\sum_{j=0}^{\infty} T_j$ of distributions will be said to converge in $\mathscr{D}'$ and to have distribution S as its sum if the sequence of partial sums $S_k = \sum_{j=0}^{k}$ converges to S.

These definitions of convergence in $\mathscr{D}'$ assume that the limits T and S are known in advance, and are distributions. This raises the question of the *completeness* of $\mathscr{D}'$: if a sequence (T_j) in $\mathscr{D}'$ is such that the sequence $(\langle T_j, \varphi\rangle)$ has a limit in $\mathbb{C}$ for all $\varphi \in \mathscr{D}$, does the map

$$\varphi \longmapsto \lim_{j\to\infty} \langle T_j, \varphi\rangle$$

define a distribution $T \in \mathscr{D}'$? In other words, does the limiting process preserve continuity with respect to φ? It is a remarkable theorem that, because of the strong topology on $\mathscr{D}$, this is actually the case. An analogous statement holds for series. This notion of convergence does not coincide with any of the classical notions used for ordinary functions: for example, the sequence (φ_ν) with $\varphi_\nu(x) = \cos \nu x$ converges to 0 in $\mathscr{D}'(\mathbb{R})$, but fails to do so by any of the standard criteria.

An example of convergent sequences of distributions is provided by sequences which converge to δ. If (f_ν) is a sequence of locally summable functions on $\mathbb{R}^n$ such that

(i) $\int_{\|\mathbf{x}\|<b} f_\nu(\mathbf{x})\,\mathrm{d}^n\mathbf{x} \to 1$ as $\nu \to \infty$ for all $b > 0$;

(ii) $\int_{a\le\|\mathbf{x}\|\le 1/a} |f_\nu(\mathbf{x})|\,\mathrm{d}^n\mathbf{x} \to 0$ as $\nu \to \infty$ for all $0 < a < 1$;

(iii) there exists $d > 0$ and $M > 0$ such that $\int_{\|\mathbf{x}\|<d} |f_\nu(\mathbf{x})|\,\mathrm{d}^n\mathbf{x} < M$ for all ν;

then the sequence (T_{f_ν}) of distributions converges to δ in $\mathscr{D}'(\mathbb{R}^n)$.

1.3.2.3.9. Operations on distributions

As a general rule, the definitions are chosen so that the operations coincide with those on functions whenever a distribution is associated to a function.

Most definitions consist in transferring to a distribution T an operation which is well defined on $\varphi \in \mathscr{D}$ by 'transposing' it in the duality product $\langle T, \varphi\rangle$; this procedure will map T to a new distribution provided the original operation maps $\mathscr{D}$ *continuously* into itself.

1.3.2.3.9.1. Differentiation

(*a*) *Definition and elementary properties*

If T is a distribution on $\mathbb{R}^n$, its partial derivative $\partial_i T$ with respect to x_i is defined by

$$\langle \partial_i T, \varphi\rangle = -\langle T, \partial_i \varphi\rangle$$

for all $\varphi \in \mathscr{D}$. This does define a distribution, because the partial differentiations $\varphi \longmapsto \partial_i\varphi$ are continuous for the topology of $\mathscr{D}$.

Suppose that $T = T_f$ with f a locally integrable function such that $\partial_i f$ exists and is almost everywhere continuous. Then integration by parts along the x_i axis gives

$$\begin{aligned}\int_{\mathbb{R}^n} &\partial_i f(x_1, \ldots, x_i, \ldots, x_n)\varphi(x_1, \ldots, x_i, \ldots, x_n)\,\mathrm{d}x_i \\ &= (f\varphi)(x_1, \ldots, +\infty, \ldots, x_n) - (f\varphi)(x_1, \ldots, -\infty, \ldots, x_n) \\ &\quad - \int_{\mathbb{R}^n} f(x_1, \ldots, x_i, \ldots, x_n)\partial_i\varphi(x_1, \ldots, x_i, \ldots, x_n)\,\mathrm{d}x_i;\end{aligned}$$

the integrated term vanishes, since φ has compact support, showing that $\partial_i T_f = T_{\partial_i f}$.

The test functions $\varphi \in \mathscr{D}$ are infinitely differentiable. Therefore, transpositions like that used to define $\partial_i T$ may be repeated, so that *any distribution is infinitely differentiable*. For instance,

$$\begin{aligned}\langle \partial_{ij}^2 T, \varphi\rangle &= -\langle \partial_j T, \partial_i\varphi\rangle = \langle T, \partial_{ij}^2\varphi\rangle, \\ \langle D^{\mathbf{p}} T, \varphi\rangle &= (-1)^{|\mathbf{p}|}\langle T, D^{\mathbf{p}}\varphi\rangle, \\ \langle \Delta T, \varphi\rangle &= \langle T, \Delta\varphi\rangle,\end{aligned}$$

where Δ is the Laplacian operator. The derivatives of Dirac's δ distribution are

$$\langle D^{\mathbf{p}}\delta, \varphi\rangle = (-1)^{|\mathbf{p}|}\langle \delta, D^{\mathbf{p}}\varphi\rangle = (-1)^{|\mathbf{p}|} D^{\mathbf{p}}\varphi(\mathbf{0}).$$

It is remarkable that *differentiation is a continuous operation* for the topology on $\mathscr{D}'$: if a sequence (T_j) of distributions converges to distribution T, then the sequence $(D^{\mathbf{p}} T_j)$ of derivatives converges to $D^{\mathbf{p}} T$ for any multi-index $\mathbf{p}$, since as $j \to \infty$

$$\langle D^{\mathbf{p}} T_j, \varphi\rangle = (-1)^{|\mathbf{p}|}\langle T_j, D^{\mathbf{p}}\varphi\rangle \to (-1)^{|\mathbf{p}|}\langle T, D^{\mathbf{p}}\varphi\rangle = \langle D^{\mathbf{p}} T, \varphi\rangle.$$

An analogous statement holds for series: any convergent series of distributions may be differentiated termwise to all orders. This illustrates how 'robust' the constructs of distribution theory are in comparison with those of ordinary function theory, where similar statements are notoriously untrue.

(*b*) *Differentiation under the duality bracket*

Limiting processes and differentiation may also be carried out under the duality bracket $\langle , \rangle$ as under the integral sign with ordinary functions. Let the function $\varphi = \varphi(\mathbf{x}, \lambda)$ depend on a parameter $\lambda \in \Lambda$ and a vector $\mathbf{x} \in \mathbb{R}^n$ in such a way that all functions

$$\varphi_\lambda : \mathbf{x} \longmapsto \varphi(\mathbf{x}, \lambda)$$

be in $\mathscr{D}(\mathbb{R}^n)$ for all $\lambda \in \Lambda$. Let $T \in \mathscr{D}'(\mathbb{R}^n)$ be a distribution, let

$$I(\lambda) = \langle T, \varphi_\lambda\rangle$$

and let $\lambda_0 \in \Lambda$ be given parameter value. Suppose that, as λ runs through a small enough neighbourhood of λ_0,

(i) all the φ_λ have their supports in a fixed compact subset K of $\mathbb{R}^n$;

(ii) all the derivatives $D^{\mathbf{p}}\varphi_\lambda$ have a partial derivative with respect to λ which is continuous with respect to $\mathbf{x}$ and λ.

Under these hypotheses, $I(\lambda)$ is differentiable (in the usual sense) with respect to λ near λ_0, and its derivative may be obtained by 'differentiation under the $\langle , \rangle$ sign':

$$\frac{\mathrm{d}I}{\mathrm{d}\lambda} = \langle T, \partial_\lambda\varphi_\lambda\rangle.$$

(*c*) *Effect of discontinuities*

When a function f or its derivatives are no longer continuous, the derivatives $D^{\mathbf{p}} T_f$ of the associated distribution T_f may no longer coincide with the distributions associated to the functions $D^{\mathbf{p}} f$.

In dimension 1, the simplest example is Heaviside's unit step function Y [$Y(x) = 0$ for $x < 0$, $Y(x) = 1$ for $x \ge 0$]:

$$\langle (T_Y)', \varphi\rangle = -\langle (T_Y), \varphi'\rangle = -\int_0^{+\infty} \varphi'(x)\,\mathrm{d}x = \varphi(0) = \langle \delta, \varphi\rangle.$$

Hence $(T_Y)' = \delta$, a result long used 'heuristically' by electrical engineers [see also Dirac (1958)].

Let f be infinitely differentiable for $x<0$ and $x>0$ but have discontinuous derivatives $f^{(m)}$ at $x=0$ [$f^{(0)}$ being f itself] with jumps $\sigma_m = f^{(m)}(0+) - f^{(m)}(0-)$. Consider the functions:

$$\begin{aligned} g_0 &= f - \sigma_0 Y \\ g_1 &= g_0' - \sigma_1 Y \\ &\text{- - - - - - - - - -} \\ g_k &= g_{k-1}' - \sigma_k Y. \end{aligned}$$

The g_k are continuous, their derivatives g_k' are continuous almost everywhere [which implies that $(T_{g_k})' = T_{g_k'}$ and $g_k' = f^{(k+1)}$ almost everywhere]. This yields immediately:

$$\begin{aligned} (T_f)' &= T_{f'} + \sigma_0 \delta \\ (T_f)'' &= T_{f''} + \sigma_0 \delta' + \sigma_1 \delta \\ &\text{- - - - - - - - - - - - - - - - - - -} \\ (T_f)^{(m)} &= T_{f^{(m)}} + \sigma_0 \delta^{(m-1)} + \ldots + \sigma_{m-1}\delta. \\ &\text{- - - - - - - - - - - - - - - - - - -} \end{aligned}$$

Thus the 'distributional derivatives' $(T_f)^{(m)}$ differ from the usual functional derivatives $T_{f^{(m)}}$ by singular terms associated with discontinuities.

In dimension n, let f be infinitely differentiable everywhere except on a smooth hypersurface S, across which its partial derivatives show discontinuities. Let σ_0 and σ_ν denote the discontinuities of f and its normal derivative $\partial_\nu \varphi$ across S (both σ_0 and σ_ν are functions of position on S), and let $\delta_{(S)}$ and $\partial_\nu \delta_{(S)}$ be defined by

$$\begin{aligned} \langle \delta_{(S)}, \varphi \rangle &= \int_S \varphi \, \mathrm{d}^{n-1} S \\ \langle \partial_\nu \delta_{(S)}, \varphi \rangle &= -\int_S \partial_\nu \varphi \, \mathrm{d}^{n-1} S. \end{aligned}$$

Integration by parts shows that

$$\partial_i T_f = T_{\partial_i f} + \sigma_0 \cos\theta_i \delta_{(S)},$$

where θ_i is the angle between the x_i axis and the normal to S along which the jump σ_0 occurs, and that the Laplacian of T_f is given by

$$\Delta(T_f) = T_{\Delta f} + \sigma_\nu \delta_{(S)} + \partial_\nu [\sigma_0 \delta_{(S)}].$$

The latter result is a statement of Green's theorem in terms of distributions. It will be used in Section 1.3.4.4.3.5 to calculate the Fourier transform of the indicator function of a molecular envelope.

1.3.2.3.9.2. Integration of distributions in dimension 1

The reverse operation from differentiation, namely calculating the 'indefinite integral' of a distribution S, consists in finding a distribution T such that $T' = S$.

For all $\chi \in \mathscr{D}$ such that $\chi = \psi'$ with $\psi \in \mathscr{D}$, we must have

$$\langle T, \chi \rangle = -\langle S, \psi \rangle.$$

This condition defines T in a 'hyperplane' $\mathscr{H}$ of $\mathscr{D}$, whose equation

$$\langle 1, \chi \rangle \equiv \langle 1, \psi' \rangle = 0$$

reflects the fact that ψ has compact support.

To specify T in the whole of $\mathscr{D}$, it suffices to specify the value of $\langle T, \varphi_0 \rangle$ where $\varphi_0 \in \mathscr{D}$ is such that $\langle 1, \varphi_0 \rangle = 1$: then any $\varphi \in \mathscr{D}$ may be written uniquely as

$$\varphi = \lambda \varphi_0 + \psi'$$

with

$$\lambda = \langle 1, \varphi \rangle, \qquad \chi = \varphi - \lambda \varphi_0, \qquad \psi(x) = \int_0^x \chi(t)\, \mathrm{d}t,$$

and T is defined by

$$\langle T, \varphi \rangle = \lambda \langle T, \varphi_0 \rangle - \langle S, \psi \rangle.$$

The freedom in the choice of φ_0 means that T is defined up to an additive constant.

1.3.2.3.9.3. Multiplication of distributions by functions

The product αT of a distribution T on $\mathbb{R}^n$ by a function α over $\mathbb{R}^n$ will be defined by transposition:

$$\langle \alpha T, \varphi \rangle = \langle T, \alpha \varphi \rangle \quad \text{for all } \varphi \in \mathscr{D}.$$

In order that αT be a distribution, the mapping $\varphi \longmapsto \alpha\varphi$ must send $\mathscr{D}(\mathbb{R}^n)$ continuously into itself; hence *the multipliers α must be infinitely differentiable*. The product of two general distributions cannot be defined. The need for a careful treatment of multipliers of distributions will become clear when it is later shown (Section 1.3.2.5.8) that the Fourier transformation turns convolutions into multiplications and *vice versa*.

If T is a distribution of order m, then α needs only have continuous derivatives up to order m. For instance, δ is a distribution of order zero, and $\alpha\delta = \alpha(\mathbf{0})\delta$ is a distribution provided α is continuous; this relation is of fundamental importance in the theory of sampling and of the properties of the Fourier transformation related to sampling (Sections 1.3.2.6.4, 1.3.2.6.6). More generally, $D^{\mathbf{p}}\delta$ is a distribution of order $|\mathbf{p}|$, and the following formula holds for all $\alpha \in \mathscr{D}^{(m)}$ with $m = |\mathbf{p}|$:

$$\alpha(D^{\mathbf{p}}\delta) = \sum_{\mathbf{q} \le \mathbf{p}} (-1)^{|\mathbf{p}-\mathbf{q}|} \binom{\mathbf{p}}{\mathbf{q}} (D^{\mathbf{p}-\mathbf{q}}\alpha)(\mathbf{0}) D^{\mathbf{q}}\delta.$$

The derivative of a product is easily shown to be

$$\partial_i(\alpha T) = (\partial_i \alpha) T + \alpha(\partial_i T)$$

and generally for any multi-index $\mathbf{p}$

$$D^{\mathbf{p}}(\alpha T) = \sum_{\mathbf{q} \le \mathbf{p}} \binom{\mathbf{p}}{\mathbf{q}} (D^{\mathbf{p}-\mathbf{q}}\alpha)(\mathbf{0}) D^{\mathbf{q}} T.$$

1.3.2.3.9.4. Division of distributions by functions

Given a distribution S on $\mathbb{R}^n$ and an infinitely differentiable multiplier function α, the division problem consists in finding a distribution T such that $\alpha T = S$.

If α never vanishes, $T = S/\alpha$ is the unique answer. If $n = 1$, and if α has only isolated zeros of finite order, it can be reduced to a collection of cases where the multiplier is x^m, for which the general solution can be shown to be of the form

$$T = U + \sum_{i=0}^{m-1} c_i \delta^{(i)},$$

where U is a particular solution of the division problem $x^m U = S$ and the c_i are arbitrary constants.

In dimension $n > 1$, the problem is much more difficult, but is of fundamental importance in the theory of linear partial differential equations, since the Fourier transformation turns the problem of solving these into a division problem for distributions [see Hörmander (1963)].

1.3.2.3.9.5. Transformation of coordinates

Let σ be a smooth nonsingular change of variables in $\mathbb{R}^n$, *i.e.* an infinitely differentiable mapping from an open subset Ω of $\mathbb{R}^n$ to Ω' in $\mathbb{R}^n$, whose Jacobian

$$J(\sigma) = \det\left[\frac{\partial \sigma(\mathbf{x})}{\partial \mathbf{x}}\right]$$

vanishes nowhere in Ω. By the implicit function theorem, the inverse mapping σ^{-1} from Ω' to Ω is well defined.

If f is a locally summable function on Ω, then the function $\sigma^{\#} f$ defined by

$$(\sigma^{\#} f)(\mathbf{x}) = f[\sigma^{-1}(\mathbf{x})]$$

is a locally summable function on Ω', and for any $\varphi \in \mathscr{D}(\Omega')$ we may write:

$$\begin{aligned}\int_{\Omega'} (\sigma^{\#} f)(\mathbf{x})\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x} &= \int_{\Omega'} f[\sigma^{-1}(\mathbf{x})]\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x} \\ &= \int_{\Omega'} f(\mathbf{y})\varphi[\sigma(\mathbf{y})]|J(\sigma)|\,\mathrm{d}^n\mathbf{y} \quad \text{by } \mathbf{x} = \sigma(\mathbf{y}).\end{aligned}$$

In terms of the associated distributions

$$\langle T_{\sigma^{\#} f}, \varphi\rangle = \langle T_f, |J(\sigma)|(\sigma^{-1})^{\#}\varphi\rangle.$$

This operation can be extended to an arbitrary distribution T by defining its *image* $\sigma^{\#} T$ under coordinate transformation σ through

$$\langle \sigma^{\#} T, \varphi\rangle = \langle T, |J(\sigma)|(\sigma^{-1})^{\#}\varphi\rangle,$$

which is well defined provided that σ is *proper*, *i.e.* that $\sigma^{-1}(K)$ is compact whenever K is compact.

For instance, if $\sigma : \mathbf{x} \longmapsto \mathbf{x} + \mathbf{a}$ is a *translation* by a vector $\mathbf{a}$ in $\mathbb{R}^n$, then $|J(\sigma)| = 1$; $\sigma^{\#}$ is denoted by $\tau_{\mathbf{a}}$, and the translate $\tau_{\mathbf{a}} T$ of a distribution T is defined by

$$\langle \tau_{\mathbf{a}} T, \varphi\rangle = \langle T, \tau_{-\mathbf{a}}\varphi\rangle.$$

Let $A : \mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$ be a linear transformation defined by a nonsingular matrix $\mathbf{A}$. Then $J(A) = \det \mathbf{A}$, and

$$\langle A^{\#} T, \varphi\rangle = |\det \mathbf{A}|\langle T, (A^{-1})^{\#}\varphi\rangle.$$

This formula will be shown later (Sections 1.3.2.6.5, 1.3.4.2.1.1) to be the basis for the definition of the reciprocal lattice.

In particular, if $\mathbf{A} = -\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, A is an inversion through a centre of symmetry at the origin, and denoting $A^{\#}\varphi$ by $\check{\varphi}$ we have:

$$\langle \check{T}, \varphi\rangle = \langle T, \check{\varphi}\rangle.$$

T is called an even distribution if $\check{T} = T$, an odd distribution if $\check{T} = -T$.

If $\mathbf{A} = \lambda\mathbf{I}$ with $\lambda > 0$, A is called a *dilation* and

$$\langle A^{\#} T, \varphi\rangle = \lambda^n \langle T, (A^{-1})^{\#}\varphi\rangle.$$

Writing symbolically δ as $\delta(\mathbf{x})$ and $A^{\#}\delta$ as $\delta(\mathbf{x}/\lambda)$, we have:

$$\delta(\mathbf{x}/\lambda) = \lambda^n \delta(\mathbf{x}).$$

If $n = 1$ and f is a function with isolated simple zeros x_j, then in the same symbolic notation

$$\delta[f(x)] = \sum_j \frac{1}{|f'(x_j)|}\delta(x_j),$$

where each $\lambda_j = 1/|f'(x_j)|$ is analogous to a 'Lorentz factor' at zero x_j.

1.3.2.3.9.6. Tensor product of distributions

The purpose of this construction is to extend Fubini's theorem to distributions. Following Section 1.3.2.2.5, we may define the tensor product $L^1_{\mathrm{loc}}(\mathbb{R}^m) \otimes L^1_{\mathrm{loc}}(\mathbb{R}^n)$ as the vector space of finite linear combinations of functions of the form

$$f \otimes g : (\mathbf{x}, \mathbf{y}) \longmapsto f(\mathbf{x})g(\mathbf{y}),$$

where $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$, $f \in L^1_{\mathrm{loc}}(\mathbb{R}^m)$ and $g \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

Let $S_{\mathbf{x}}$ and $T_{\mathbf{y}}$ denote the distributions associated to f and g, respectively, the subscripts $\mathbf{x}$ and $\mathbf{y}$ acting as mnemonics for $\mathbb{R}^m$ and $\mathbb{R}^n$. It follows from Fubini's theorem (Section 1.3.2.2.5) that $f \otimes g \in L^1_{\mathrm{loc}}(\mathbb{R}^m \times \mathbb{R}^n)$, and hence defines a distribution over $\mathbb{R}^m \times \mathbb{R}^n$; the rearrangement of integral signs gives

$$\langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle = \langle S_{\mathbf{x}}, \langle T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle\rangle = \langle T_{\mathbf{y}}, \langle S_{\mathbf{x}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle\rangle$$

for all $\varphi_{\mathbf{x},\mathbf{y}} \in \mathscr{D}(\mathbb{R}^m \times \mathbb{R}^n)$. In particular, if $\varphi(\mathbf{x}, \mathbf{y}) = u(\mathbf{x})v(\mathbf{y})$ with $u \in \mathscr{D}(\mathbb{R}^m)$, $v \in \mathscr{D}(\mathbb{R}^n)$, then

$$\langle S \otimes T, u \otimes v\rangle = \langle S, u\rangle\langle T, v\rangle.$$

This construction can be extended to general distributions $S \in \mathscr{D}'(\mathbb{R}^m)$ and $T \in \mathscr{D}'(\mathbb{R}^n)$. Given any test function $\varphi \in \mathscr{D}(\mathbb{R}^m \times \mathbb{R}^n)$, let $\varphi_{\mathbf{x}}$ denote the map $\mathbf{y} \longmapsto \varphi(\mathbf{x}, \mathbf{y})$; let $\varphi_{\mathbf{y}}$ denote the map $\mathbf{x} \longmapsto \varphi(\mathbf{x}, \mathbf{y})$; and define the two functions $\theta(\mathbf{x}) = \langle T, \varphi_{\mathbf{x}}\rangle$ and $\omega(\mathbf{y}) = \langle S, \varphi_{\mathbf{y}}\rangle$. Then, by the lemma on differentiation under the $\langle , \rangle$ sign of Section 1.3.2.3.9.1, $\theta \in \mathscr{D}(\mathbb{R}^m)$, $\omega \in \mathscr{D}(\mathbb{R}^n)$, and there exists a unique distribution $S \otimes T$ such that

$$\langle S \otimes T, \varphi \rangle = \langle S, \theta \rangle = \langle T, \omega \rangle.$$

$S \otimes T$ is called the *tensor product* of S and T.

With the mnemonic introduced above, this definition reads identically to that given above for distributions associated to locally integrable functions:

$$\langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}} \rangle = \langle S_{\mathbf{x}}, \langle T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}} \rangle \rangle = \langle T_{\mathbf{y}}, \langle S_{\mathbf{x}}, \varphi_{\mathbf{x},\mathbf{y}} \rangle \rangle.$$

The tensor product of distributions is associative:

$$(R \otimes S) \otimes T = R \otimes (S \otimes T).$$

Derivatives may be calculated by

$$D^{\mathbf{p}}_{\mathbf{x}} D^{\mathbf{q}}_{\mathbf{y}} (S_{\mathbf{x}} \otimes T_{\mathbf{y}}) = (D^{\mathbf{p}}_{\mathbf{x}} S_{\mathbf{x}}) \otimes (D^{\mathbf{q}}_{\mathbf{y}} T_{\mathbf{y}}).$$

The support of a tensor product is the Cartesian product of the supports of the two factors.

1.3.2.3.9.7. Convolution of distributions

The convolution $f * g$ of two functions f and g on $\mathbb{R}^n$ is defined by

$$(f * g)(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{y}) g(\mathbf{x} - \mathbf{y}) \, \mathrm{d}^n\mathbf{y} = \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) \, \mathrm{d}^n\mathbf{y}$$

whenever the integral exists. This is the case when f and g are both in $L^1(\mathbb{R}^n)$; then $f * g$ is also in $L^1(\mathbb{R}^n)$. Let S, T and W denote the distributions associated to f, g and $f * g$, respectively: a change of variable immediately shows that for any $\varphi \in \mathscr{D}(\mathbb{R}^n)$,

$$\langle W, \varphi \rangle = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(\mathbf{x}) g(\mathbf{y}) \varphi(\mathbf{x} + \mathbf{y}) \, \mathrm{d}^n\mathbf{x} \, \mathrm{d}^n\mathbf{y}.$$

Introducing the map σ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$ defined by $\sigma(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \mathbf{y}$, the latter expression may be written:

$$\langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi \circ \sigma \rangle$$

(where $\circ$ denotes the composition of mappings) or by a slight abuse of notation:

$$\langle W, \varphi \rangle = \langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi(\mathbf{x} + \mathbf{y}) \rangle.$$

A difficulty arises in extending this definition to general distributions S and T because the mapping σ is not proper: if K is compact in $\mathbb{R}^n$, then $\sigma^{-1}(K)$ is a cylinder with base K and generator the 'second bisector' $\mathbf{x} + \mathbf{y} = \mathbf{0}$ in $\mathbb{R}^n \times \mathbb{R}^n$. However, $\langle S \otimes T, \varphi \circ \sigma \rangle$ is defined whenever the intersection between Supp $(S \otimes T) =$ (Supp S) $\times$ (Supp T) and σ^{-1}(Supp φ) is compact.

We may therefore define the *convolution* $S * T$ of two distributions S and T on $\mathbb{R}^n$ by

$$\langle S * T, \varphi \rangle = \langle S \otimes T, \varphi \circ \sigma \rangle = \langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi(\mathbf{x} + \mathbf{y}) \rangle$$

whenever the following *support condition* is fulfilled:

> 'the set $\{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \in A, \mathbf{y} \in B, \mathbf{x} + \mathbf{y} \in K\}$ is compact in $\mathbb{R}^n \times \mathbb{R}^n$ for all K compact in $\mathbb{R}^n$'.

The latter condition is met, in particular, if S or T has compact support. The support of $S * T$ is easily seen to be contained in the closure of the vector sum

$$A + B = \{\mathbf{x} + \mathbf{y} | \mathbf{x} \in A, \mathbf{y} \in B\}.$$

Convolution by a fixed distribution S is a *continuous* operation for the topology on $\mathscr{D}'$: it maps convergent sequences (T_j) to convergent sequences $(S * T_j)$. Convolution is commutative: $S * T = T * S$.

The convolution of p distributions $T_1, \ldots, T_p$ with supports $A_1, \ldots, A_p$ can be defined by

$$\langle T_1 * \ldots * T_p, \varphi \rangle = \langle (T_1)_{\mathbf{x}_1} \otimes \ldots \otimes (T_p)_{\mathbf{x}_p}, \varphi(\mathbf{x}_1 + \ldots + \mathbf{x}_p) \rangle$$

whenever the following generalized support condition:

> 'the set $\{(\mathbf{x}_1, \ldots, \mathbf{x}_p) | \mathbf{x}_1 \in A_1, \ldots, \mathbf{x}_p \in A_p, \mathbf{x}_1 + \ldots + \mathbf{x}_p \in K\}$ is compact in $(\mathbb{R}^n)^p$ for all K compact in $\mathbb{R}^n$'

is satisfied. It is then associative. Interesting examples of associativity failure, which can be traced back to violations of the support condition, may be found in Bracewell (1986, pp. 436–437).

It follows from previous definitions that, for all distributions $T \in \mathscr{D}'$, the following identities hold:

(i) $\delta * T = T$: δ is the unit convolution;

(ii) $\delta_{(\mathbf{a})} * T = \tau_{\mathbf{a}} T$: translation is a convolution with the corresponding translate of δ;

(iii) $(D^{\mathbf{p}}\delta) * T = D^{\mathbf{p}} T$: differentiation is a convolution with the corresponding derivative of δ;

(iv) translates or derivatives of a convolution may be obtained by translating or differentiating any one of the factors: convolution 'commutes' with translation and differentiation, a property used in Section 1.3.4.4.7.7 to speed up least-squares model refinement for macromolecules.

The latter property is frequently used for the purpose of *regularization*: if T is a distribution, α an infinitely differentiable function, and at least one of the two has compact support, then $T * \alpha$ is an infinitely differentiable ordinary function. Since sequences (α_ν) of such functions α can be constructed which have compact support and converge to δ, it follows that any distribution T can be obtained as the limit of infinitely differentiable functions $T * \alpha_\nu$. In topological jargon: $\mathscr{D}(\mathbb{R}^n)$ is 'everywhere dense' in $\mathscr{D}'(\mathbb{R}^n)$. A standard function in $\mathscr{D}$ which is often used for such proofs is defined as follows: put

$$\theta(x) = \frac{1}{A} \exp\left(-\frac{1}{1 - x^2}\right) \quad \text{for } |x| \leq 1,$$

$$= 0 \qquad\qquad\qquad\qquad \text{for } |x| \geq 1,$$

with

$$A = \int_{-1}^{+1} \exp\left(-\frac{1}{1 - x^2}\right) \mathrm{d}x$$

(so that θ is in $\mathscr{D}$ and is normalized), and put

$$\theta_\varepsilon(x) = \frac{1}{\varepsilon} \theta\left(\frac{x}{\varepsilon}\right) \quad \text{in dimension 1,}$$

$$\theta_\varepsilon(\mathbf{x}) = \prod_{j=1}^{n} \theta_\varepsilon(x_j) \quad \text{in dimension } n.$$

Another related result, also proved by convolution, is the *structure theorem*: the restriction of a distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ to a bounded open set Ω in $\mathbb{R}^n$ is a derivative of finite order of a continuous function.

Properties (i) to (iv) are the basis of the symbolic or operational calculus (see Carslaw & Jaeger, 1948; Van der Pol & Bremmer, 1955; Churchill, 1958; Erdélyi, 1962; Moore, 1971) for solving integro-differential equations with constant coefficients by turning them into convolution equations, then using factorization methods for convolution algebras (Schwartz, 1965).

1.3.2.4. Fourier transforms of functions

1.3.2.4.1. Introduction

Given a complex-valued function f on $\mathbb{R}^n$ subject to suitable regularity conditions, its Fourier transform $\mathscr{F}[f]$ and Fourier cotransform $\bar{\mathscr{F}}[f]$ are defined as follows:

$$\mathscr{F}[f](\xi) = \int_{\mathbb{R}^n} f(\mathbf{x}) \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{x})\, \mathrm{d}^n\mathbf{x}$$
$$\bar{\mathscr{F}}[f](\xi) = \int_{\mathbb{R}^n} f(\mathbf{x}) \exp(+2\pi i \boldsymbol{\xi}\cdot\mathbf{x})\, \mathrm{d}^n\mathbf{x},$$

where $\boldsymbol{\xi}\cdot\mathbf{x} = \sum_{i=1}^n \xi_i x_i$ is the ordinary scalar product. The terminology and sign conventions given above are the standard ones in mathematics; those used in crystallography are slightly different (see Section 1.3.4.2.1.1). These transforms enjoy a number of remarkable properties, whose natural settings entail different regularity assumptions on f: for instance, properties relating to convolution are best treated in $L^1(\mathbb{R}^n)$, while Parseval's theorem requires the Hilbert space structure of $L^2(\mathbb{R}^n)$. After a brief review of these classical properties, the Fourier transformation will be examined in a space $\mathscr{S}(\mathbb{R}^n)$ particularly well suited to accommodating the full range of its properties, which will later serve as a space of test functions to extend the Fourier transformation to distributions.

There exists an abundant literature on the 'Fourier integral'. The books by Carslaw (1930), Wiener (1933), Titchmarsh (1948), Katznelson (1968), Sneddon (1951, 1972), and Dym & McKean (1972) are particularly recommended.

1.3.2.4.2. Fourier transforms in L^1

1.3.2.4.2.1. Linearity

Both transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ are obviously linear maps from L^1 to L^∞ when these spaces are viewed as vector spaces over the field $\mathbb{C}$ of complex numbers.

1.3.2.4.2.2. Effect of affine coordinate transformations

$\mathscr{F}$ and $\bar{\mathscr{F}}$ turn translations into phase shifts:

$$\mathscr{F}[\tau_{\mathbf{a}} f](\xi) = \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{a})\mathscr{F}[f](\xi)$$
$$\bar{\mathscr{F}}[\tau_{\mathbf{a}} f](\xi) = \exp(+2\pi i \boldsymbol{\xi}\cdot\mathbf{a})\bar{\mathscr{F}}[f](\xi).$$

Under a general linear change of variable $\mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$ with nonsingular matrix $\mathbf{A}$, the transform of $A^{\#}f$ is

$$\begin{aligned}\mathscr{F}[A^{\#}f](\xi) &= \int_{\mathbb{R}^n} f(\mathbf{A}^{-1}\mathbf{x}) \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{x})\, \mathrm{d}^n\mathbf{x} \\ &= \int_{\mathbb{R}^n} f(\mathbf{y}) \exp(-2\pi i (A^T\boldsymbol{\xi})\cdot\mathbf{y}) |\det\mathbf{A}|\, \mathrm{d}^n\mathbf{y} \\ &\qquad\qquad \text{by } \mathbf{x} = \mathbf{A}\mathbf{y} \\ &= |\det\mathbf{A}| \mathscr{F}[f](\mathbf{A}^T\boldsymbol{\xi})\end{aligned}$$

i.e.

$$\mathscr{F}[A^{\#}f] = |\det\mathbf{A}|[(\mathbf{A}^{-1})^T]^{\#}\mathscr{F}[f]$$

and similarly for $\bar{\mathscr{F}}$. The matrix $(\mathbf{A}^{-1})^T$ is called the *contragredient* of matrix $\mathbf{A}$.

Under an affine change of coordinates $\mathbf{x} \longmapsto S(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ with nonsingular matrix $\mathbf{A}$, the transform of $S^{\#}f$ is given by

$$\begin{aligned}\mathscr{F}[S^{\#}f](\xi) &= \mathscr{F}[\tau_{\mathbf{b}}(A^{\#}f)](\xi) \\ &= \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{b})\mathscr{F}[A^{\#}f](\xi) \\ &= \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{b})|\det\mathbf{A}|\mathscr{F}[f](\mathbf{A}^T\boldsymbol{\xi})\end{aligned}$$

with a similar result for $\bar{\mathscr{F}}$, replacing $-i$ by $+i$.

1.3.2.4.2.3. Conjugate symmetry

The kernels of the Fourier transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ satisfy the following identities:

$$\exp(\pm 2\pi i \boldsymbol{\xi}\cdot\mathbf{x}) = \exp\overline{[\pm 2\pi i \boldsymbol{\xi}\cdot(-\mathbf{x})]} = \exp\overline{[\pm 2\pi i(-\boldsymbol{\xi})\cdot\mathbf{x}]}.$$

As a result the transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ themselves have the following 'conjugate symmetry' properties [where the notation $\check{f}(\mathbf{x}) = f(-\mathbf{x})$ of Section 1.3.2.2.2 will be used]:

$$\mathscr{F}[f](\xi) = \overline{\mathscr{F}[\bar{f}](-\xi)} = \overline{\mathscr{F}[\check{\bar{f}}](\xi)}$$
$$\mathscr{F}[f](\xi) = \overline{\mathscr{F}[\check{\bar{f}}](\xi)}.$$

Therefore,

(i) f real $\Leftrightarrow f = \bar{f} \Leftrightarrow \mathscr{F}[f] = \overline{\check{\mathscr{F}}[f]} \Leftrightarrow \mathscr{F}[f](\xi) = \overline{\mathscr{F}[f](-\xi)}$: $\mathscr{F}[f]$ is said to possess *Hermitian symmetry*;

(ii) f centrosymmetric $\Leftrightarrow f = \check{f} \Leftrightarrow \mathscr{F}[f] = \overline{\mathscr{F}[\bar{f}]}$;

(iii) f real centrosymmetric $\Leftrightarrow f = \bar{f} = \check{f} \Leftrightarrow \mathscr{F}[f] = \overline{\mathscr{F}[f]} = \overline{\check{\mathscr{F}}[f]} \Leftrightarrow \mathscr{F}[f]$ real centrosymmetric.

Conjugate symmetry is the basis of Friedel's law (Section 1.3.4.2.1.4) in crystallography.

1.3.2.4.2.4. Tensor product property

Another elementary property of $\mathscr{F}$ is its naturality with respect to tensor products. Let $u \in L^1(\mathbb{R}^m)$ and $v \in L^1(\mathbb{R}^n)$, and let $\mathscr{F}_{\mathbf{x}}, \mathscr{F}_{\mathbf{y}}, \mathscr{F}_{\mathbf{x},\mathbf{y}}$ denote the Fourier transformations in $L^1(\mathbb{R}^m)$, $L^1(\mathbb{R}^n)$ and $L^1(\mathbb{R}^m \times \mathbb{R}^n)$, respectively. Then

$$\mathscr{F}_{\mathbf{x},\mathbf{y}}[u \otimes v] = \mathscr{F}_{\mathbf{x}}[u] \otimes \mathscr{F}_{\mathbf{y}}[v].$$

Furthermore, if $f \in L^1(\mathbb{R}^m \times \mathbb{R}^n)$, then $\mathscr{F}_{\mathbf{y}}[f] \in L^1(\mathbb{R}^m)$ as a function of $\mathbf{x}$ and $\mathscr{F}_{\mathbf{x}}[f] \in L^1(\mathbb{R}^n)$ as a function of $\mathbf{y}$, and

$$\mathscr{F}_{\mathbf{x},\mathbf{y}}[f] = \mathscr{F}_{\mathbf{x}}[\mathscr{F}_{\mathbf{y}}[f]] = \mathscr{F}_{\mathbf{y}}[\mathscr{F}_{\mathbf{x}}[f]].$$

This is easily proved by using Fubini's theorem and the fact that $(\boldsymbol{\xi},\boldsymbol{\eta})\cdot(\mathbf{x},\mathbf{y}) = \boldsymbol{\xi}\cdot\mathbf{x} + \boldsymbol{\eta}\cdot\mathbf{y}$, where $\mathbf{x},\boldsymbol{\xi} \in \mathbb{R}^m$, $\mathbf{y},\boldsymbol{\eta} \in \mathbb{R}^n$. This property may be written:

$$\mathscr{F}_{\mathbf{x},\mathbf{y}} = \mathscr{F}_{\mathbf{x}} \otimes \mathscr{F}_{\mathbf{y}}.$$

1.3.2.4.2.5. Convolution property

If f and g are summable, their convolution $f * g$ exists and is summable, and

$$\mathscr{F}[f*g](\boldsymbol{\xi}) = \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x}-\mathbf{y})\,\mathrm{d}^n\mathbf{y}\right]\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\mathbf{x}.$$

With $\mathbf{x} = \mathbf{y} + \mathbf{z}$, so that

$$\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x}) = \exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{y})\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{z}),$$

and with Fubini's theorem, rearrangement of the double integral gives:

$$\mathscr{F}[f*g] = \mathscr{F}[f]\times\mathscr{F}[g]$$

and similarly

$$\bar{\mathscr{F}}[f*g] = \bar{\mathscr{F}}[f]\times\bar{\mathscr{F}}[g].$$

Thus *the Fourier transform and cotransform turn convolution into multiplication*.

1.3.2.4.2.6. Reciprocity property

In general, $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are not summable, and hence cannot be further transformed; however, as they are essentially bounded, their products with the Gaussians $G_t(\xi) = \exp(-2\pi^2\|\xi\|^2 t)$ are summable for all $t > 0$, and it can be shown that

$$f = \lim_{t\to 0}\bar{\mathscr{F}}[G_t\mathscr{F}[f]] = \lim_{t\to 0}\mathscr{F}[G_t\bar{\mathscr{F}}[f]],$$

where the limit is taken in the topology of the L^1 norm $\|.\|_1$. Thus $\mathscr{F}$ and $\bar{\mathscr{F}}$ are (in a sense) mutually inverse, which justifies the common practice of calling $\bar{\mathscr{F}}$ the 'inverse Fourier transformation'.

1.3.2.4.2.7. Riemann–Lebesgue lemma

If $f \in L^1(\mathbb{R}^n)$, *i.e.* is summable, then $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ exist and are continuous and essentially bounded:

$$\|\mathscr{F}[f]\|_\infty = \|\bar{\mathscr{F}}[f]\|_\infty \le \|f\|_1.$$

In fact one has the much stronger property, whose statement constitutes the *Riemann–Lebesgue lemma*, that $\mathscr{F}[f](\boldsymbol{\xi})$ and $\bar{\mathscr{F}}[f](\boldsymbol{\xi})$ both tend to zero as $\|\boldsymbol{\xi}\| \to \infty$.

1.3.2.4.2.8. Differentiation

Let us now suppose that $n = 1$ and that $f \in L^1(\mathbb{R})$ is differentiable with $f' \in L^1(\mathbb{R})$. Integration by parts yields

$$\begin{aligned}\mathscr{F}[f'](\xi) &= \int_{-\infty}^{+\infty} f'(x)\exp(-2\pi i\xi\cdot x)\,\mathrm{d}x\\ &= [f(x)\exp(-2\pi i\xi\cdot x)]_{-\infty}^{+\infty}\\ &\quad + 2\pi i\xi\int_{-\infty}^{+\infty} f(x)\exp(-2\pi i\xi\cdot x)\,\mathrm{d}x.\end{aligned}$$

Since f' is summable, f has a limit when $x \to \pm\infty$, and this limit must be 0 since f is summable. Therefore

$$\mathscr{F}[f'](\xi) = (2\pi i\xi)\mathscr{F}[f](\xi)$$

with the bound

$$\|2\pi\xi\mathscr{F}[f]\|_\infty \le \|f'\|_1$$

so that $|\mathscr{F}[f](\xi)|$ decreases faster than $1/|\xi| \to \infty$.

This result can be easily extended to several dimensions and to any multi-index $\mathbf{m}$: if f is summable and has continuous summable partial derivatives up to order $|\mathbf{m}|$, then

$$\mathscr{F}[D^{\mathbf{m}}f](\boldsymbol{\xi}) = (2\pi i\boldsymbol{\xi})^{\mathbf{m}}\mathscr{F}[f](\boldsymbol{\xi})$$

and

$$\|(2\pi\boldsymbol{\xi})^{\mathbf{m}}\mathscr{F}[f]\|_\infty \le \|D^{\mathbf{m}}f\|_1.$$

Similar results hold for $\bar{\mathscr{F}}$, with $2\pi i\boldsymbol{\xi}$ replaced by $-2\pi i\boldsymbol{\xi}$. Thus, the more differentiable f is, with summable derivatives, the faster $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ decrease at infinity.

The property of turning differentiation into multiplication by a monomial has many important applications in crystallography, for instance differential syntheses (Sections 1.3.4.2.1.9, 1.3.4.4.7.2, 1.3.4.4.7.5) and moment-generating functions [Section 1.3.4.5.2.1(*c*)].

1.3.2.4.2.9. Decrease at infinity

Conversely, assume that f is summable on $\mathbb{R}^n$ and that f decreases fast enough at infinity for $\mathbf{x}^{\mathbf{m}}f$ also to be summable, for some multi-index $\mathbf{m}$. Then the integral defining $\mathscr{F}[f]$ may be subjected to the differential operator $D^{\mathbf{m}}$, still yielding a convergent integral: therefore $D^{\mathbf{m}}\mathscr{F}[f]$ exists, and

$$D^{\mathbf{m}}(\mathscr{F}[f])(\boldsymbol{\xi}) = \mathscr{F}[(-2\pi i\mathbf{x})^{\mathbf{m}}f](\boldsymbol{\xi})$$

with the bound

$$\|D^{\mathbf{m}}(\mathscr{F}[f])\|_\infty = \|(2\pi\mathbf{x})^{\mathbf{m}}f\|_1.$$

Similar results hold for $\bar{\mathscr{F}}$, with $-2\pi i\mathbf{x}$ replaced by $2\pi i\mathbf{x}$. Thus, the faster f decreases at infinity, the more $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are differentiable, with bounded derivatives. This property is the converse of that described in Section 1.3.2.4.2.8, and their combination is fundamental in the definition of the function space $\mathscr{S}$ in Section 1.3.2.4.4.1, of tempered distributions in Section 1.3.2.5, and in the extension of the Fourier transformation to them.

1.3.2.4.2.10. The Paley–Wiener theorem

An extreme case of the last instance occurs when f has *compact support*: then $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are so regular that they may be analytically continued from $\mathbb{R}^n$ to $\mathbb{C}^n$ where they are *entire* functions, *i.e.* have no singularities at finite distance (Paley & Wiener, 1934). This is easily seen for $\mathscr{F}[f]$: giving vector $\boldsymbol{\xi} \in \mathbb{R}^n$ a vector $\boldsymbol{\eta} \in \mathbb{R}^n$ of imaginary parts leads to

$$\begin{aligned}\mathscr{F}[f](\boldsymbol{\xi}+i\boldsymbol{\eta}) &= \int_{\mathbb{R}^n} f(\mathbf{x})\exp[-2\pi i(\boldsymbol{\xi}+i\boldsymbol{\eta})\cdot\mathbf{x}]\,\mathrm{d}^n\mathbf{x}\\ &= \mathscr{F}[\exp(2\pi\boldsymbol{\eta}\cdot\mathbf{x})f](\boldsymbol{\xi}),\end{aligned}$$

where the latter transform always exists since $\exp(2\pi\boldsymbol{\eta}\cdot\mathbf{x})f$ is summable with respect to $\mathbf{x}$ for all values of $\boldsymbol{\eta}$. This analytic continuation forms the basis of the saddlepoint method in probability theory [Section 1.3.4.5.2.1(*f*)] and leads to the use of maximum-entropy distributions in the statistical theory of direct phase determination [Section 1.3.4.5.2.2(*e*)].

By Liouville's theorem, an entire function in $\mathbb{C}^n$ cannot vanish identically on the complement of a compact subset of $\mathbb{R}^n$ without vanishing everywhere: therefore $\mathscr{F}[f]$ cannot have compact support if f has, and hence $\mathscr{D}(\mathbb{R}^n)$ *is not stable by Fourier transformation*.

1.3.2.4.3. Fourier transforms in L^2

Let f belong to $L^2(\mathbb{R}^n)$, *i.e.* be such that

$$\|f\|_2 = \left(\int_{\mathbb{R}^n} |f(\mathbf{x})|^2 \,\mathrm{d}^n\mathbf{x}\right)^{1/2} < \infty.$$

1.3.2.4.3.1. Invariance of L^2

$\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ exist and are functions in L^2, *i.e.* $\mathscr{F}L^2 = L^2$, $\bar{\mathscr{F}}L^2 = L^2$.

1.3.2.4.3.2. Reciprocity

$\mathscr{F}[\bar{\mathscr{F}}[f]] = f$ and $\bar{\mathscr{F}}[\mathscr{F}[f]] = f$, equality being taken as 'almost everywhere' equality. This again leads to calling $\bar{\mathscr{F}}$ the 'inverse Fourier transformation' rather than the Fourier cotransformation.

1.3.2.4.3.3. Isometry

$\mathscr{F}$ and $\bar{\mathscr{F}}$ preserve the L^2 norm:

$\|\mathscr{F}[f]\|_2 = \|\bar{\mathscr{F}}[f]\|_2 = \|f\|_2$ (Parseval's/Plancherel's theorem).

This property, which may be written in terms of the inner product $(,)$ in $L^2(\mathbb{R}^n)$ as

$$(\mathscr{F}[f], \mathscr{F}[g]) = (\bar{\mathscr{F}}[f], \bar{\mathscr{F}}[g]) = (f, g),$$

implies that $\mathscr{F}$ and $\bar{\mathscr{F}}$ are *unitary* transformations of $L^2(\mathbb{R}^n)$ into itself, *i.e.* infinite-dimensional 'rotations'.

1.3.2.4.3.4. Eigenspace decomposition of L^2

Some light can be shed on the geometric structure of these rotations by the following simple considerations. Note that

$$\begin{aligned}\mathscr{F}^2[f](\mathbf{x}) &= \int_{\mathbb{R}^n} \mathscr{F}[f](\boldsymbol{\xi}) \exp(-2\pi i \mathbf{x}\cdot\boldsymbol{\xi}) \,\mathrm{d}^n\boldsymbol{\xi} \\ &= \bar{\mathscr{F}}[\mathscr{F}[f]](-\mathbf{x}) = f(-\mathbf{x})\end{aligned}$$

so that $\mathscr{F}^4$ (and similarly $\bar{\mathscr{F}}^4$) is the identity map. Any eigenvalue of $\mathscr{F}$ or $\bar{\mathscr{F}}$ is therefore a fourth root of unity, *i.e.* ± 1 or $\pm i$, and $L^2(\mathbb{R}^n)$ splits into an orthogonal direct sum

$$\mathbf{H}_0 \otimes \mathbf{H}_1 \otimes \mathbf{H}_2 \otimes \mathbf{H}_3,$$

where $\mathscr{F}$ (respectively $\bar{\mathscr{F}}$) acts in each subspace $\mathbf{H}_k$ $(k = 0, 1, 2, 3)$ by multiplication by $(-i)^k$. Orthonormal bases for these subspaces can be constructed from Hermite functions (*cf.* Section 1.3.2.4.4.2) This method was used by Wiener (1933, pp. 51–71).

1.3.2.4.3.5. The convolution theorem and the isometry property

In L^2, the convolution theorem (when applicable) and the Parseval/Plancherel theorem are not independent. Suppose that f, g, $f \times g$ and $f * g$ are all in L^2 (without questioning whether these properties are independent). Then $f * g$ may be written in terms of the inner product in L^2 as follows:

$$(f * g)(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y}) g(\mathbf{y}) \,\mathrm{d}^n\mathbf{y} = \int_{\mathbb{R}^n} \overline{\check{\bar{f}}(\mathbf{y}-\mathbf{x})} g(\mathbf{y}) \,\mathrm{d}^n\mathbf{y},$$

i.e.

$$(f * g)(\mathbf{x}) = (\tau_{\mathbf{x}} \check{\bar{f}}, g).$$

Invoking the isometry property, we may rewrite the right-hand side as

$$\begin{aligned}(\mathscr{F}[\tau_{\mathbf{x}} \check{\bar{f}}], \mathscr{F}[g]) &= (\exp(-2\pi i \mathbf{x}\cdot\boldsymbol{\xi}) \overline{\mathscr{F}[f]}_{\xi}, \mathscr{F}[g]_{\xi}) \\ &= \int_{\mathbb{R}^n} (\mathscr{F}[f] \times \mathscr{F}[g])(\mathbf{x}) \\ &\quad \times \exp(+2\pi i \mathbf{x}\cdot\boldsymbol{\xi}) \,\mathrm{d}^n\boldsymbol{\xi} \\ &= \bar{\mathscr{F}}[\mathscr{F}[f] \times \mathscr{F}[g]],\end{aligned}$$

so that the initial identity yields the convolution theorem.

To obtain the converse implication, note that

$$\begin{aligned}(f, g) &= \int_{\mathbb{R}^n} \overline{f(\mathbf{y})} g(\mathbf{y}) \,\mathrm{d}^n\mathbf{y} = (\check{\bar{f}} * g)(\mathbf{0}) \\ &= \bar{\mathscr{F}}[\mathscr{F}[\check{\bar{f}}] \times \mathscr{F}[g]](\mathbf{0}) \\ &= \int_{\mathbb{R}^n} \overline{\mathscr{F}[f](\boldsymbol{\xi})} \mathscr{F}[g](\boldsymbol{\xi}) \,\mathrm{d}^n\boldsymbol{\xi} = (\mathscr{F}[f], \mathscr{F}[g]),\end{aligned}$$

where conjugate symmetry (Section 1.3.2.4.2.2) has been used.

These relations have an important application in the calculation by Fourier transform methods of the derivatives used in the refinement of macromolecular structures (Section 1.3.4.4.7).

1.3.2.4.4. Fourier transforms in $\mathscr{S}$

1.3.2.4.4.1. Definition and properties of $\mathscr{S}$

The duality established in Sections 1.3.2.4.2.8 and 1.3.2.4.2.9 between the local differentiability of a function and the rate of decrease at infinity of its Fourier transform prompts one to consider the space $\mathscr{S}(\mathbb{R}^n)$ of functions f on $\mathbb{R}^n$ which are infinitely differentiable *and* all of whose derivatives are rapidly decreasing, so that for all multi-indices $\mathbf{k}$ and $\mathbf{p}$

$$(\mathbf{x}^{\mathbf{k}} D^{\mathbf{p}} f)(\mathbf{x}) \to 0 \quad \text{as} \quad \|\mathbf{x}\| \to \infty.$$

The product of $f \in \mathscr{S}$ by any polynomial over $\mathbb{R}^n$ is still in $\mathscr{S}$ ($\mathscr{S}$ is an algebra over the ring of polynomials). Furthermore, $\mathscr{S}$ is invariant under translations and differentiation.

If $f \in \mathscr{S}$, then its transforms $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are

(i) infinitely differentiable because f is rapidly decreasing;

(ii) rapidly decreasing because f is infinitely differentiable;

hence $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are in $\mathscr{S}$: $\mathscr{S}$ *is invariant* under $\mathscr{F}$ and $\bar{\mathscr{F}}$.

Since $L^1 \supset \mathscr{S}$ and $L^2 \supset \mathscr{S}$, all properties of $\mathscr{F}$ and $\bar{\mathscr{F}}$ already encountered above are enjoyed by functions of $\mathscr{S}$, with all restrictions on differentiability and/or integrability lifted. For instance, given two functions f and g in $\mathscr{S}$, then both fg and $f * g$ are in $\mathscr{S}$ (which was not the case with L^1 nor with L^2) so that the reciprocity theorem inherited from L^2

$$\mathscr{F}[\bar{\mathscr{F}}[f]] = f \quad \text{and} \quad \bar{\mathscr{F}}[\mathscr{F}[f]] = f$$

allows one to state the reverse of the convolution theorem first established in L^1:

$$\mathscr{F}[fg] = \mathscr{F}[f] * \mathscr{F}[g]$$
$$\bar{\mathscr{F}}[fg] = \bar{\mathscr{F}}[f] * \bar{\mathscr{F}}[g].$$

1.3.2.4.4.2. Gaussian functions and Hermite functions

Gaussian functions are particularly important elements of $\mathscr{S}$. In dimension 1, a well known contour integration (Schwartz, 1965, p. 184) yields

$$\mathscr{F}[\exp(-\pi x^2)](\xi) = \bar{\mathscr{F}}[\exp(-\pi x^2)](\xi) = \exp(-\pi\xi^2),$$

which shows that the 'standard Gaussian' $\exp(-\pi x^2)$ is invariant under $\mathscr{F}$ and $\bar{\mathscr{F}}$. By a tensor product construction, it follows that the same is true of the standard Gaussian

$$G(\mathbf{x}) = \exp(-\pi\|\mathbf{x}\|^2)$$

in dimension n:

$$\mathscr{F}[G](\boldsymbol{\xi}) = \bar{\mathscr{F}}[G](\boldsymbol{\xi}) = G(\boldsymbol{\xi}).$$

In other words, G is an eigenfunction of $\mathscr{F}$ and $\bar{\mathscr{F}}$ for eigenvalue 1 (Section 1.3.2.4.3.4).

A complete system of eigenfunctions may be constructed as follows. In dimension 1, consider the family of functions

$$H_m = \frac{D^m G^2}{G} \quad (m \geq 0),$$

where D denotes the differentiation operator. The first two members of the family

$$H_0 = G, \qquad H_1 = 2DG,$$

are such that $\mathscr{F}[H_0] = H_0$, as shown above, and

$$DG(x) = -2\pi x G(x) = i(2\pi i x)G(x) = i\mathscr{F}[DG](x),$$

hence

$$\mathscr{F}[H_1] = (-i)H_1.$$

We may thus take as an induction hypothesis that

$$\mathscr{F}[H_m] = (-i)^m H_m.$$

The identity

$$D\left(\frac{D^m G^2}{G}\right) = \frac{D^{m+1}G^2}{G} - \frac{DG}{G}\frac{D^m G^2}{G}$$

may be written

$$H_{m+1}(x) = (DH_m)(x) - 2\pi x H_m(x),$$

and the two differentiation theorems give:

$$\mathscr{F}[DH_m](\xi) = (2\pi i\xi)\mathscr{F}[H_m](\xi)$$
$$\mathscr{F}[-2\pi x H_m](\xi) = -iD(\mathscr{F}[H_m])(\xi).$$

Combination of this with the induction hypothesis yields

$$\mathscr{F}[H_{m+1}](\xi) = (-i)^{m+1}[(DH_m)(\xi) - 2\pi\xi H_m(\xi)]$$
$$= (-i)^{m+1}H_{m+1}(\xi),$$

thus proving that H_m is an eigenfunction of $\mathscr{F}$ for eigenvalue $(-i)^m$ for all $m \geq 0$. The same proof holds for $\bar{\mathscr{F}}$, with eigenvalue i^m. If these eigenfunctions are normalized as

$$\mathscr{H}_m(x) = \frac{(-1)^m 2^{1/4}}{\sqrt{m!}2^m\pi^{m/2}}H_m(x),$$

then it can be shown that the collection of *Hermite functions* $\{\mathscr{H}_m(x)\}_{m\geq 0}$ constitutes an orthonormal basis of $L^2(\mathbb{R})$ such that $\mathscr{H}_m$ is an eigenfunction of $\mathscr{F}$ (respectively $\bar{\mathscr{F}}$) for eigenvalue $(-i)^m$ (respectively i^m).

In dimension n, the same construction can be extended by tensor product to yield the multivariate Hermite functions

$$\mathscr{H}_{\mathbf{m}}(\mathbf{x}) = \mathscr{H}_{m_1}(x_1) \times \mathscr{H}_{m_2}(x_2) \times \ldots \times \mathscr{H}_{m_n}(x_n)$$

(where $\mathbf{m} \geq \mathbf{0}$ is a multi-index). These constitute an orthonormal basis of $L^2(\mathbb{R}^n)$, with $\mathscr{H}_{\mathbf{m}}$ an eigenfunction of $\mathscr{F}$ (respectively $\bar{\mathscr{F}}$) for eigenvalue $(-i)^{|\mathbf{m}|}$ (respectively $i^{|\mathbf{m}|}$). Thus the subspaces $\mathbf{H}_k$ of Section 1.3.2.4.3.4 are spanned by those $\mathscr{H}_{\mathbf{m}}$ with $|\mathbf{m}| \equiv k \bmod 4$ ($k = 0, 1, 2, 3$).

General multivariate Gaussians are usually encountered in the nonstandard form

$$G_{\mathbf{A}}(\mathbf{x}) = \exp(-\tfrac{1}{2}\mathbf{x}^T \cdot \mathbf{A}\mathbf{x}),$$

where $\mathbf{A}$ is a symmetric positive-definite matrix. Diagonalizing $\mathbf{A}$ as $\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^T$ with $\mathbf{E}\mathbf{E}^T$ the identity matrix, and putting $\mathbf{A}^{1/2} = \mathbf{E}\boldsymbol{\Lambda}^{1/2}\mathbf{E}^T$, we may write

$$G_{\mathbf{A}}(\mathbf{x}) = G\left[\left(\frac{\mathbf{A}}{2\pi}\right)^{1/2}\mathbf{x}\right]$$

i.e.

$$G_{\mathbf{A}} = [(2\pi\mathbf{A}^{-1})^{1/2}]^{\#}G;$$

hence (by Section 1.3.2.4.2.3)

$$\mathscr{F}[G_{\mathbf{A}}] = |\det(2\pi\mathbf{A}^{-1})|^{1/2}\left[\left(\frac{\mathbf{A}}{2\pi}\right)^{1/2}\right]^{\#}G,$$

i.e.

$$\mathscr{F}[G_{\mathbf{A}}](\boldsymbol{\xi}) = |\det(2\pi\mathbf{A}^{-1})|^{1/2}G[(2\pi\mathbf{A}^{-1})^{1/2}\boldsymbol{\xi}],$$

i.e. finally

$$\mathscr{F}[G_{\mathbf{A}}] = |\det(2\pi\mathbf{A}^{-1})|^{1/2} G_{4\pi^2\mathbf{A}^{-1}}.$$

This result is widely used in crystallography, *e.g.* to calculate form factors for anisotropic atoms (Section 1.3.4.2.2.6) and to obtain transforms of derivatives of Gaussian atomic densities (Section 1.3.4.4.7.10).

1.3.2.4.4.3. Heisenberg's inequality, Hardy's theorem

The result just obtained, which also holds for $\bar{\mathscr{F}}$, shows that the 'peakier' $G_{\mathbf{A}}$, the 'broader' $\mathscr{F}[G_{\mathbf{A}}]$. This is a general property of the Fourier transformation, expressed in dimension 1 by the *Heisenberg inequality* (Weyl, 1931):

$$\left(\int x^2 |f(x)|^2 \,\mathrm{d}x\right)\left(\int \xi^2 |\mathscr{F}[f](\xi)|^2 \,\mathrm{d}\xi\right) \geq \frac{1}{16\pi^2}\left(\int |f(x)|^2 \,\mathrm{d}x\right)^2,$$

where, by a beautiful theorem of Hardy (1933), equality can only be attained for f Gaussian. Hardy's theorem is even stronger: if both f and $\mathscr{F}[f]$ behave at infinity as constant multiples of G, then each of them is *everywhere* a constant multiple of G; if both f and $\mathscr{F}[f]$ behave at infinity as constant multiples of $G \times$ monomial, then each of them is a finite linear combination of Hermite functions. Hardy's theorem is invoked in Section 1.3.4.4.5 to derive the optimal procedure for spreading atoms on a sampling grid in order to obtain the most accurate structure factors.

The search for optimal compromises between the confinement of f to a compact domain in x-space and of $\mathscr{F}[f]$ to a compact domain in ξ-space leads to consideration of prolate spheroidal wavefunctions (Pollack & Slepian, 1961; Landau & Pollack, 1961, 1962).

1.3.2.4.4.4. Symmetry property

A final formal property of the Fourier transform, best established in $\mathscr{S}$, is its *symmetry*: if f and g are in $\mathscr{S}$, then by Fubini's theorem

$$\begin{aligned}\langle \mathscr{F}[f], g\rangle &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(\mathbf{x}) \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{x}) \,\mathrm{d}^n\mathbf{x}\right) g(\boldsymbol{\xi}) \,\mathrm{d}^n\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^n} f(\mathbf{x})\left(\int_{\mathbb{R}^n} g(\boldsymbol{\xi}) \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{x}) \,\mathrm{d}^n\boldsymbol{\xi}\right) \mathrm{d}^n\mathbf{x} \\ &= \langle f, \mathscr{F}[g]\rangle.\end{aligned}$$

This possibility of 'transposing' $\mathscr{F}$ (and $\bar{\mathscr{F}}$) from the left to the right of the duality bracket will be used in Section 1.3.2.5.4 to extend the Fourier transformation to distributions.

1.3.2.4.5. Various writings of Fourier transforms

Other ways of writing Fourier transforms in $\mathbb{R}^n$ exist besides the one used here. All have the form

$$\mathscr{F}_{h,\omega}[f](\boldsymbol{\xi}) = \frac{1}{h^n}\int_{\mathbb{R}^n} f(\mathbf{x}) \exp(-i\omega\boldsymbol{\xi}\cdot\mathbf{x}) \,\mathrm{d}^n\mathbf{x},$$

where h is real positive and ω real nonzero, with the reciprocity formula written:

$$f(\mathbf{x}) = \frac{1}{k^n}\int_{\mathbb{R}^n} \mathscr{F}_{h,\omega}[f](\boldsymbol{\xi}) \exp(+i\omega\boldsymbol{\xi}\cdot\mathbf{x}) \,\mathrm{d}^n\mathbf{x}$$

with k real positive. The consistency condition between h, k and ω is

$$hk = \frac{2\pi}{|\omega|}.$$

The usual choices are:

(i) $\omega = \pm 2\pi, h = k = 1$ (as here);
(ii) $\omega = \pm 1, h = 1, k = 2\pi$ (in probability theory and in solid-state physics);
(iii) $\omega = \pm 1, h = k = \sqrt{2\pi}$ (in much of classical analysis).

It should be noted that conventions (ii) and (iii) introduce numerical factors of 2π in convolution and Parseval formulae, while (ii) breaks the symmetry between $\mathscr{F}$ and $\bar{\mathscr{F}}$.

1.3.2.4.6. Tables of Fourier transforms

The books by Campbell & Foster (1948), Erdélyi (1954) and Magnus *et al.* (1966) contain extensive tables listing pairs of functions and their Fourier transforms. Bracewell (1986) lists those pairs particularly relevant to electrical engineering applications.

1.3.2.5. Fourier transforms of tempered distributions

1.3.2.5.1. Introduction

It was found in Section 1.3.2.4.2 that the usual space of test functions $\mathscr{D}$ is not invariant under $\mathscr{F}$ and $\bar{\mathscr{F}}$. By contrast, the space $\mathscr{S}$ of infinitely differentiable rapidly decreasing functions *is* invariant under $\mathscr{F}$ and $\bar{\mathscr{F}}$, and furthermore transposition formulae such as

$$\langle \mathscr{F}[f], g\rangle = \langle f, \mathscr{F}[g]\rangle$$

hold for all $f, g \in \mathscr{S}$. It is precisely this type of transposition which was used successfully in Sections 1.3.2.3.9.1 and 1.3.2.3.9.3 to define the derivatives of distributions and their products with smooth functions.

This suggests using $\mathscr{S}$ instead of $\mathscr{D}$ as a space of test functions φ, and defining the Fourier transform $\mathscr{F}[T]$ of a distribution T by

$$\langle \mathscr{F}[T], \varphi\rangle = \langle T, \mathscr{F}[\varphi]\rangle$$

whenever T is capable of being extended from $\mathscr{D}$ to $\mathscr{S}$ while remaining continuous. It is this latter proviso which will be subsumed under the adjective 'tempered'. As was the case with the construction of $\mathscr{D}'$, it is the definition of a sufficiently strong topology (*i.e.* notion of convergence) in $\mathscr{S}$ which will play a key role in transferring to the elements of its topological dual $\mathscr{S}'$ (called tempered distributions) all the properties of the Fourier transformation.

Besides the general references to distribution theory mentioned in Section 1.3.2.3.1 the reader may consult the books by Zemanian (1965, 1968). Lavoine (1963) contains tables of Fourier transforms of distributions.

1.3.2.5.2. $\mathscr{S}$ as a test-function space

A notion of convergence has to be introduced in $\mathscr{S}(\mathbb{R}^n)$ in order to be able to define and test the continuity of linear functionals on it.

A sequence (φ_j) of functions in $\mathscr{S}$ will be said to converge to 0 if, for any given multi-indices $\mathbf{k}$ and $\mathbf{p}$, the sequence $(\mathbf{x}^{\mathbf{k}} D^{\mathbf{p}} \varphi_j)$ tends to 0 uniformly on $\mathbb{R}^n$.

It can be shown that $\mathscr{D}(\mathbb{R}^n)$ is dense in $\mathscr{S}(\mathbb{R}^n)$. Translation is continuous for this topology. For any linear differential operator $P(D) = \sum_{\mathbf{p}} a_{\mathbf{p}} D^{\mathbf{p}}$ and any polynomial $Q(\mathbf{x})$ over $\mathbb{R}^n$, $(\varphi_j) \to 0$ implies $[Q(\mathbf{x}) \times P(D)\varphi_j] \to 0$ in the topology of $\mathscr{S}$. Therefore, differentiation and multiplication by polynomials are continuous for the topology on $\mathscr{S}$.

The Fourier transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ are also continuous for the topology of $\mathscr{S}$. Indeed, let (φ_j) converge to 0 for the topology on $\mathscr{S}$. Then, by Section 1.3.2.4.2,

$$\|(2\pi\boldsymbol{\xi})^{\mathbf{m}} D^{\mathbf{p}}(\mathscr{F}[\varphi_j])\|_\infty \leq \|D^{\mathbf{m}}[(2\pi\mathbf{x})^{\mathbf{p}} \varphi_j]\|_1.$$

The right-hand side tends to 0 as $j \to \infty$ by definition of convergence in $\mathscr{S}$, hence $\|\boldsymbol{\xi}\|^{\mathbf{m}} D^{\mathbf{p}}(\mathscr{F}[\varphi_j]) \to 0$ uniformly, so that $(\mathscr{F}[\varphi_j]) \to 0$ in $\mathscr{S}$ as $j \to \infty$. The same proof applies to $\bar{\mathscr{F}}$.

1.3.2.5.3. Definition and examples of tempered distributions

A distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ is said to be *tempered* if it can be extended into a continuous linear functional on $\mathscr{S}$.

If $\mathscr{S}'(\mathbb{R}^n)$ is the topological dual of $\mathscr{S}(\mathbb{R}^n)$, and if $S \in \mathscr{S}'(\mathbb{R}^n)$, then its restriction to $\mathscr{D}$ is a tempered distribution; conversely, if $T \in \mathscr{D}'$ is tempered, then its extension to $\mathscr{S}$ is unique (because $\mathscr{D}$ is dense in $\mathscr{S}$), hence it defines an element S of $\mathscr{S}'$. We may therefore identify $\mathscr{S}'$ and the space of tempered distributions.

A distribution with compact support is tempered, *i.e.* $\mathscr{S}' \supset \mathscr{E}'$. By transposition of the corresponding properties of $\mathscr{S}$, it is readily established that the derivative, translate or product by a polynomial of a tempered distribution is still a tempered distribution.

These inclusion relations may be summarized as follows: since $\mathscr{S}$ contains $\mathscr{D}$ but is contained in $\mathscr{E}$, the reverse inclusions hold for the topological duals, and hence $\mathscr{S}'$ contains $\mathscr{E}'$ but is contained in $\mathscr{D}'$.

A locally summable function f on $\mathbb{R}^n$ will be said to be *of polynomial growth* if $|f(\mathbf{x})|$ can be majorized by a polynomial in $\|\mathbf{x}\|$ as $\|\mathbf{x}\| \to \infty$. It is easily shown that such a function f defines a tempered distribution T_f *via*

$$\langle T_f, \varphi \rangle = \int_{\mathbb{R}^n} f(\mathbf{x}) \varphi(\mathbf{x}) \, \mathrm{d}^n \mathbf{x}.$$

In particular, polynomials over $\mathbb{R}^n$ define tempered distributions, and so do functions in $\mathscr{S}$. The latter remark, together with the transposition identity (Section 1.3.2.4.4), invites the extension of $\mathscr{F}$ and $\bar{\mathscr{F}}$ from $\mathscr{S}$ to $\mathscr{S}'$.

1.3.2.5.4. Fourier transforms of tempered distributions

The Fourier transform $\mathscr{F}[T]$ and cotransform $\bar{\mathscr{F}}[T]$ of a tempered distribution T are defined by

$$\langle \mathscr{F}[T], \varphi \rangle = \langle T, \mathscr{F}[\varphi] \rangle$$
$$\langle \bar{\mathscr{F}}[T], \varphi \rangle = \langle T, \bar{\mathscr{F}}[\varphi] \rangle$$

for all test functions $\varphi \in \mathscr{S}$. Both $\mathscr{F}[T]$ and $\bar{\mathscr{F}}[T]$ are themselves tempered distributions, since the maps $\varphi \longmapsto \mathscr{F}[\varphi]$ and $\varphi \longmapsto \bar{\mathscr{F}}[\varphi]$ are both linear and continuous for the topology of $\mathscr{S}$. In the same way that $\mathbf{x}$ and $\boldsymbol{\xi}$ have been used consistently as arguments for φ and $\mathscr{F}[\varphi]$, respectively, the notation $T_{\mathbf{x}}$ and $\mathscr{F}[T]_{\boldsymbol{\xi}}$ will be used to indicate which variables are involved.

When T is a distribution with compact support, its Fourier transform may be written

$$\mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}} = \langle T_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \rangle$$

since the function $\mathbf{x} \longmapsto \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x})$ is in $\mathscr{E}$ while $T_{\mathbf{x}} \in \mathscr{E}'$. It can be shown, as in Section 1.3.2.4.2, to be analytically continuable into an entire function over $\mathbb{C}^n$.

1.3.2.5.5. Transposition of basic properties

The duality between differentiation and multiplication by a monomial extends from $\mathscr{S}$ to $\mathscr{S}'$ by transposition:

$$\mathscr{F}[D_{\mathbf{x}}^{\mathbf{p}} T_{\mathbf{x}}]_{\boldsymbol{\xi}} = (2\pi i \boldsymbol{\xi})^{\mathbf{p}} \mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}$$
$$D_{\boldsymbol{\xi}}^{\mathbf{p}}(\mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}) = \mathscr{F}[(-2\pi i \mathbf{x})^{\mathbf{p}} T_{\mathbf{x}}]_{\boldsymbol{\xi}}.$$

Analogous formulae hold for $\bar{\mathscr{F}}$, with i replaced by $-i$.

The formulae expressing the duality between translation and phase shift, *e.g.*

$$\mathscr{F}[\tau_{\mathbf{a}} T_{\mathbf{x}}]_{\boldsymbol{\xi}} = \exp(-2\pi i \mathbf{a} \cdot \boldsymbol{\xi}) \mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}$$
$$\tau_{\boldsymbol{a}}(\mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}) = \mathscr{F}[\exp(2\pi i \boldsymbol{a} \cdot \mathbf{x}) T_{\mathbf{x}}]_{\boldsymbol{\xi}};$$

between a linear change of variable and its contragredient, *e.g.*

$$\mathscr{F}[A^{\#} T] = |\det \mathbf{A}| [(\mathbf{A}^{-1})^T]^{\#} \mathscr{F}[T];$$

are obtained similarly by transposition from the corresponding identities in $\mathscr{S}$. They give a transposition formula for an affine change of variables $\mathbf{x} \longmapsto S(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ with nonsingular matrix $\mathbf{A}$:

$$\begin{aligned} \mathscr{F}[S^{\#} T] &= \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{b}) \mathscr{F}[A^{\#} T] \\ &= \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{b}) |\det \mathbf{A}| [(\mathbf{A}^{-1})^T]^{\#} \mathscr{F}[T], \end{aligned}$$

with a similar result for $\bar{\mathscr{F}}$, replacing $-i$ by $+i$.

Conjugate symmetry is obtained similarly:

$$\mathscr{F}[\bar{T}] = \overline{\check{\mathscr{F}}[T]}, \; \mathscr{F}[\check{\bar{T}}] = \overline{\mathscr{F}[T]},$$

with the same identities for $\bar{\mathscr{F}}$.

The tensor product property also transposes to tempered distributions: if $U \in \mathscr{S}'(\mathbb{R}^m)$, $V \in \mathscr{S}'(\mathbb{R}^n)$,

$$\mathscr{F}[U_{\mathbf{x}} \otimes V_{\mathbf{y}}] = \mathscr{F}[U]_{\boldsymbol{\xi}} \otimes \mathscr{F}[V]_{\boldsymbol{\eta}}$$
$$\bar{\mathscr{F}}[U_{\mathbf{x}} \otimes V_{\mathbf{y}}] = \bar{\mathscr{F}}[U]_{\boldsymbol{\xi}} \otimes \bar{\mathscr{F}}[V]_{\boldsymbol{\eta}}.$$

1.3.2.5.6. Transforms of δ-functions

Since δ has compact support,

$$\mathscr{F}[\delta_{\mathbf{x}}]_{\boldsymbol{\xi}} = \langle \delta_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \rangle = 1_{\boldsymbol{\xi}}, \quad \textit{i.e.} \; \mathscr{F}[\delta] = 1.$$

It is instructive to show that conversely $\mathscr{F}[1] = \delta$ without invoking the reciprocity theorem. Since $\partial_j 1 = 0$ for all $j = 1, \ldots, n$, it follows from Section 1.3.2.3.9.4 that $\mathscr{F}[1] = c\delta$; the

constant c can be determined by using the invariance of the standard Gaussian G established in Section 1.3.2.4.3:

$$\langle \mathscr{F}[1]_{\mathbf{x}}, G_{\mathbf{x}} \rangle = \langle 1_{\xi}, G_{\xi} \rangle = 1;$$

hence $c = 1$. Thus, $\mathscr{F}[1] = \delta$.

The basic properties above then read (using multi-indices to denote differentiation):

$$\mathscr{F}[\delta_{\mathbf{x}}^{(\mathbf{m})}]_{\xi} = (2\pi i \boldsymbol{\xi})^{\mathbf{m}}, \quad \mathscr{F}[\mathbf{x}^{\mathbf{m}}]_{\xi} = (-2\pi i)^{-|\mathbf{m}|} \delta_{\xi}^{(\mathbf{m})};$$
$$\mathscr{F}[\delta_{\mathbf{a}}]_{\xi} = \exp(-2\pi i \mathbf{a} \cdot \boldsymbol{\xi}), \quad \mathscr{F}[\exp(2\pi i \boldsymbol{\alpha} \cdot \mathbf{x})]_{\xi} = \delta_{\boldsymbol{\alpha}},$$

with analogous relations for $\bar{\mathscr{F}}$, i becoming $-i$. Thus derivatives of δ are mapped to monomials (and *vice versa*), while translates of δ are mapped to 'phase factors' (and *vice versa*).

1.3.2.5.7. Reciprocity theorem

The previous results now allow a self-contained and rigorous proof of the reciprocity theorem between $\mathscr{F}$ and $\bar{\mathscr{F}}$ to be given, whereas in traditional settings (*i.e.* in L^1 and L^2) the implicit handling of δ through a limiting process is always the sticking point.

Reciprocity is first established in $\mathscr{S}$ as follows:

$$\begin{aligned}\bar{\mathscr{F}}[\mathscr{F}[\varphi]](\mathbf{x}) &= \int_{\mathbb{R}^n} \mathscr{F}[\varphi](\boldsymbol{\xi}) \exp(2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \, \mathrm{d}^n \boldsymbol{\xi} \\ &= \int_{\mathbb{R}^n} \mathscr{F}[\tau_{-\mathbf{x}}\varphi](\boldsymbol{\xi}) \, \mathrm{d}^n \boldsymbol{\xi} \\ &= \langle 1, \mathscr{F}[\tau_{-\mathbf{x}}\varphi] \rangle \\ &= \langle \mathscr{F}[1], \tau_{-\mathbf{x}}\varphi \rangle \\ &= \langle \tau_{\mathbf{x}}\delta, \varphi \rangle \\ &= \varphi(\mathbf{x})\end{aligned}$$

and similarly

$$\mathscr{F}[\bar{\mathscr{F}}[\varphi]](\mathbf{x}) = \varphi(\mathbf{x}).$$

The reciprocity theorem is then proved in $\mathscr{S}'$ by transposition:

$$\bar{\mathscr{F}}[\mathscr{F}[T]] = \mathscr{F}[\bar{\mathscr{F}}[T]] = T \quad \text{for all } T \in \mathscr{S}'.$$

Thus the Fourier cotransformation $\bar{\mathscr{F}}$ in $\mathscr{S}'$ may legitimately be called the 'inverse Fourier transformation'.

The method of Section 1.3.2.4.3 may then be used to show that $\mathscr{F}$ and $\bar{\mathscr{F}}$ both have period 4 in $\mathscr{S}'$.

1.3.2.5.8. Multiplication and convolution

Multiplier functions $\alpha(\mathbf{x})$ for tempered distributions must be infinitely differentiable, as for ordinary distributions; furthermore, they must grow sufficiently slowly as $\|x\| \to \infty$ to ensure that $\alpha\varphi \in \mathscr{S}$ for all $\varphi \in \mathscr{S}$ and that the map $\varphi \longmapsto \alpha\varphi$ is continuous for the topology of $\mathscr{S}$. This leads to choosing for multipliers the subspace $\mathscr{O}_M$ consisting of functions $\alpha \in \mathscr{E}$ *of polynomial growth*. It can be shown that if f is in $\mathscr{O}_M$, then the associated distribution T_f is in $\mathscr{S}'$ (*i.e.* is a tempered distribution); and that conversely if T is in $\mathscr{S}'$, $\mu * T$ is in $\mathscr{O}_M$ for all $\mu \in \mathscr{D}$.

Corresponding restrictions must be imposed to define the space $\mathscr{O}'_C$ of those distributions T whose convolution $S * T$ with a tempered distribution S is still a tempered distribution: T must be such that, for all $\varphi \in \mathscr{S}$, $\theta(\mathbf{x}) = \langle T_{\mathbf{y}}, \varphi(\mathbf{x} + \mathbf{y}) \rangle$ is in $\mathscr{S}$; and such that the map $\varphi \longmapsto \theta$ be continuous for the topology of $\mathscr{S}$. This implies that S is 'rapidly decreasing'. It can be shown that if f is in $\mathscr{S}$, then the associated distribution T_f is in $\mathscr{O}'_C$; and that conversely if T is in $\mathscr{O}'_C$, $\mu * T$ is in $\mathscr{S}$ for all $\mu \in \mathscr{D}$.

The two spaces $\mathscr{O}_M$ and $\mathscr{O}'_C$ are mapped into each other by the Fourier transformation

$$\mathscr{F}(\mathscr{O}_M) = \bar{\mathscr{F}}(\mathscr{O}_M) = \mathscr{O}'_C$$
$$\mathscr{F}(\mathscr{O}'_C) = \bar{\mathscr{F}}(\mathscr{O}'_C) = \mathscr{O}_M$$

and the convolution theorem takes the form

$$\mathscr{F}[\alpha S] = \mathscr{F}[\alpha] * \mathscr{F}[S] \quad S \in \mathscr{S}', \alpha \in \mathscr{O}_M, \mathscr{F}[\alpha] \in \mathscr{O}'_C;$$
$$\mathscr{F}[S * T] = \mathscr{F}[S] \times \mathscr{F}[T] \quad S \in \mathscr{S}', T \in \mathscr{O}'_C, \mathscr{F}[T] \in \mathscr{O}_M.$$

The same identities hold for $\bar{\mathscr{F}}$. Taken together with the reciprocity theorem, these show that $\mathscr{F}$ and $\bar{\mathscr{F}}$ establish mutually inverse isomorphisms between $\mathscr{O}_M$ and $\mathscr{O}'_C$, and exchange multiplication for convolution in $\mathscr{S}'$.

It may be noticed that most of the basic properties of $\mathscr{F}$ and $\bar{\mathscr{F}}$ may be deduced from this theorem and from the properties of δ. Differentiation operators $D^{\mathbf{m}}$ and translation operators $\tau_{\mathbf{a}}$ are convolutions with $D^{\mathbf{m}}\delta$ and $\tau_{\mathbf{a}}\delta$; they are turned, respectively, into multiplication by monomials $(\pm 2\pi i \boldsymbol{\xi})^{\mathbf{m}}$ (the transforms of $D^{\mathbf{m}}\delta$) or by phase factors $\exp(\pm 2\pi i \boldsymbol{\xi} \cdot \boldsymbol{\alpha})$ (the transforms of $\tau_{\mathbf{a}}\delta$).

Another consequence of the convolution theorem is the duality established by the Fourier transformation between sections and projections of a function and its transform. For instance, in $\mathbb{R}^3$, the *projection* of $f(x, y, z)$ on the x, y plane along the z axis may be written

$$(\delta_x \otimes \delta_y \otimes 1_z) * f;$$

its Fourier transform is then

$$(1_{\xi} \otimes 1_{\eta} \otimes \delta_{\zeta}) \times \mathscr{F}[f],$$

which is the *section* of $\mathscr{F}[f]$ by the plane $\zeta = 0$, orthogonal to the z axis used for projection. There are numerous applications of this property in crystallography (Section 1.3.4.2.1.8) and in fibre diffraction (Section 1.3.4.5.1.3).

1.3.2.5.9. L^2 aspects, Sobolev spaces

The special properties of $\mathscr{F}$ in the space of square-integrable functions $L^2(\mathbb{R}^n)$, such as Parseval's identity, can be accommodated within distribution theory: if $u \in L^2(\mathbb{R}^n)$, then T_u is a tempered distribution in $\mathscr{S}'$ (the map $u \longmapsto T_u$ being continuous) and it can be shown that $S = \mathscr{F}[T_u]$ is of the form S_v, where $u = \mathscr{F}[u]$ is the Fourier transform of u in $L^2(\mathbb{R}^n)$. By Plancherel's theorem, $\|u\|_2 = \|v\|_2$.

This embedding of L^2 into $\mathscr{S}'$ can be used to derive the convolution theorem for L^2. If u and v are in $L^2(\mathbb{R}^n)$, then $u * v$ can be shown to be a bounded continuous function; thus $u * v$ is not in L^2, but it is in $\mathscr{S}'$, so that its Fourier transform is a distribution, and

$$\mathscr{F}[u * v] = \mathscr{F}[u] \times \mathscr{F}[v].$$

Spaces of tempered distributions related to $L^2(\mathbb{R}^n)$ can be defined as follows. For any real s, define the Sobolev space $H_s(\mathbb{R}^n)$ to consist of all tempered distributions $S \in \mathscr{S}'(\mathbb{R}^n)$ such that

$$(1 + |\boldsymbol{\xi}|^2)^{s/2} \mathscr{F}[S]_{\xi} \in L^2(\mathbb{R}^n).$$

These spaces play a fundamental role in the theory of partial differential equations, and in the mathematical theory of tomographic reconstruction – a subject not unrelated to the crystallographic phase problem (Natterer, 1986).

1.3.2.6. Periodic distributions and Fourier series

1.3.2.6.1. Terminology

Let $\mathbb{Z}^n$ be the subset of $\mathbb{R}^n$ consisting of those points with (signed) integer coordinates; it is an n-dimensional *lattice*, *i.e.* a free Abelian group on n generators. A particularly simple set of n generators is given by the standard basis of $\mathbb{R}^n$, and hence $\mathbb{Z}^n$ will be called the *standard lattice in* $\mathbb{R}^n$. Any other 'nonstandard' n-dimensional lattice Λ in $\mathbb{R}^n$ is the image of this standard lattice by a general linear transformation.

If we identify any two points in $\mathbb{R}^n$ whose coordinates are congruent modulo $\mathbb{Z}^n$, *i.e.* differ by a vector in $\mathbb{Z}^n$, we obtain the *standard n-torus* $\mathbb{R}^n/\mathbb{Z}^n$. The latter may be viewed as $(\mathbb{R}/\mathbb{Z})^n$, *i.e.* as the Cartesian product of n circles. The same identification may be carried out modulo a nonstandard lattice Λ, yielding a *nonstandard n-torus* $\mathbb{R}^n/\Lambda$. The correspondence to crystallographic terminology is that 'standard' coordinates over the standard 3-torus $\mathbb{R}^3/\mathbb{Z}^3$ are called 'fractional' coordinates over the unit cell; while Cartesian coordinates, *e.g.* in ångströms, constitute a set of nonstandard coordinates.

Finally, we will denote by I the unit cube $[0, 1]^n$ and by C_ε the subset

$$C_\varepsilon = \{\mathbf{x} \in \mathbb{R}^n \| x_j| < \varepsilon \text{ for all } j = 1, \ldots, n\}.$$

1.3.2.6.2. $\mathbb{Z}^n$-periodic distributions in $\mathbb{R}^n$

A distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ is called *periodic with period lattice* $\mathbb{Z}^n$ (or $\mathbb{Z}^n$-periodic) if $\tau_{\mathbf{m}} T = T$ for all $\mathbf{m} \in \mathbb{Z}^n$ (in crystallography the period lattice is the *direct* lattice).

Given a distribution with compact support $T^0 \in \mathscr{E}'(\mathbb{R}^n)$, then $T = \sum_{\mathbf{m}\in\mathbb{Z}^n} \tau_{\mathbf{m}} T^0$ is a $\mathbb{Z}^n$-periodic distribution. Note that we may write $T = r * T^0$, where $r = \sum_{\mathbf{m}\in\mathbb{Z}^n} \delta_{(\mathbf{m})}$ consists of Dirac δ's at all nodes of the period lattice $\mathbb{Z}^n$.

Conversely, any $\mathbb{Z}^n$-periodic distribution T may be written as $r * T^0$ for some $T^0 \in \mathscr{E}'$. To retrieve such a 'motif' T^0 from T, a function ψ will be constructed in such a way that $\psi \in \mathscr{D}$ (hence has compact support) and $r * \psi = 1$; then $T^0 = \psi T$. Indicator functions (Section 1.3.2.2) such as χ_1 or $\chi_{C_{1/2}}$ cannot be used directly, since they are discontinuous; but regularized versions of them may be constructed by convolution (see Section 1.3.2.3.9.7) as $\psi_0 = \chi_{C_\varepsilon} * \theta_\eta$, with ε and η such that $\psi_0(\mathbf{x}) = 1$ on $C_{1/2}$ and $\psi_0(\mathbf{x}) = 0$ outside $C_{3/4}$. Then the function

$$\psi = \frac{\psi_0}{\sum_{\mathbf{m}\in\mathbb{Z}^n} \tau_{\mathbf{m}} \psi_0}$$

has the desired property. The sum in the denominator contains at most 2^n nonzero terms at any given point $\mathbf{x}$ and acts as a smoothly varying 'multiplicity correction'.

1.3.2.6.3. Identification with distributions over $\mathbb{R}^n/\mathbb{Z}^n$

Throughout this section, 'periodic' will mean '$\mathbb{Z}^n$-periodic'.

Let $s \in \mathbb{R}$, and let $[s]$ denote the largest integer $\leq s$. For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let $\tilde{\mathbf{x}}$ be the unique vector $(\tilde{x}_1, \ldots, \tilde{x}_n)$ with $\tilde{x}_j = x_j - [x_j]$. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then $\tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ if and only if $\mathbf{x} - \mathbf{y} \in \mathbb{Z}^n$. The image of the map $\mathbf{x} \longmapsto \tilde{\mathbf{x}}$ is thus $\mathbb{R}^n$ modulo $\mathbb{Z}^n$, or $\mathbb{R}^n/\mathbb{Z}^n$.

If f is a periodic function over $\mathbb{R}^n$, then $\tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ implies $f(\mathbf{x}) = f(\mathbf{y})$; we may thus define a function $\tilde{f}$ over $\mathbb{R}^n/\mathbb{Z}^n$ by putting $\tilde{f}(\tilde{\mathbf{x}}) = f(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} - \tilde{\mathbf{x}} \in \mathbb{Z}^n$. Conversely, if $\tilde{f}$ is a function over $\mathbb{R}^n/\mathbb{Z}^n$, then we may define a function f over $\mathbb{R}^n$ by putting $f(\mathbf{x}) = \tilde{f}(\tilde{\mathbf{x}})$, and f will be periodic. Periodic functions over $\mathbb{R}^n$ may thus be identified with functions over $\mathbb{R}^n/\mathbb{Z}^n$, and this identification preserves the notions of convergence, local summability and differentiability.

Given $\varphi^0 \in \mathscr{D}(\mathbb{R}^n)$, we may define

$$\varphi(\mathbf{x}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} (\tau_{\mathbf{m}} \varphi^0)(\mathbf{x})$$

since the sum only contains finitely many nonzero terms; φ is periodic, and $\tilde{\varphi} \in \mathscr{D}(\mathbb{R}^n/\mathbb{Z}^n)$. Conversely, if $\tilde{\varphi} \in \mathscr{D}(\mathbb{R}^n/\mathbb{Z}^n)$ we may define $\varphi \in \mathscr{E}(\mathbb{R}^n)$ periodic by $\varphi(\mathbf{x}) = \tilde{\varphi}(\tilde{\mathbf{x}})$, and $\varphi^0 \in \mathscr{D}(\mathbb{R}^n)$ by putting $\varphi^0 = \psi\varphi$ with ψ constructed as above.

By transposition, a distribution $\tilde{T} \in \mathscr{D}'(\mathbb{R}^n/\mathbb{Z}^n)$ defines a unique periodic distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ by $\langle T, \varphi^0\rangle = \langle \tilde{T}, \tilde{\varphi}\rangle$; conversely, $T \in \mathscr{D}'(\mathbb{R}^n)$ periodic defines uniquely $\tilde{T} \in \mathscr{D}'(\mathbb{R}^n/\mathbb{Z}^n)$ by $\langle \tilde{T}, \tilde{\varphi}\rangle = \langle T, \varphi^0\rangle$.

We may therefore identify $\mathbb{Z}^n$-periodic distributions over $\mathbb{R}^n$ with distributions over $\mathbb{R}^n/\mathbb{Z}^n$. We will, however, use mostly the former presentation, as it is more closely related to the crystallographer's perception of periodicity (see Section 1.3.4.1).

1.3.2.6.4. Fourier transforms of periodic distributions

The content of this section is perhaps the central result in the relation between Fourier theory and crystallography (Section 1.3.4.2.1.1).

Let $T = r * T^0$ with r defined as in Section 1.3.2.6.2. Then $r \in \mathscr{S}'$, $T^0 \in \mathscr{E}'$ hence $T^0 \in \mathscr{O}'_C$, so that $T \in \mathscr{S}'$: $\mathbb{Z}^n$-periodic distributions are tempered, hence have a Fourier transform. The convolution theorem (Section 1.3.2.5.8) is applicable, giving:

$$\mathscr{F}[T] = \mathscr{F}[r] \times \mathscr{F}[T^0]$$

and similarly for $\bar{\mathscr{F}}$.

Since $\mathscr{F}[\delta_{(\mathbf{m})}](\xi) = \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{m})$, formally

$$\mathscr{F}[r]_\xi = \sum_{\mathbf{m}\in\mathbb{Z}^n} \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{m}) = Q,$$

say.

It is readily shown that Q is tempered and periodic, so that $Q = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \tau_{\boldsymbol{\mu}}(\psi Q)$, while the periodicity of r implies that

$$[\exp(-2\pi i \xi_j) - 1]\psi Q = 0, \quad j = 1, \ldots, n.$$

Since the first factors have single isolated zeros at $\xi_j = 0$ in $C_{3/4}$, $\psi Q = c\delta$ (see Section 1.3.2.3.9.4) and hence by periodicity $Q = cr$; convoluting with χ_{C_1} shows that $c = 1$. Thus we have the fundamental result:

$$\boxed{\mathscr{F}[r] = r}$$

so that

$$\mathscr{F}[T] = r \times \mathscr{F}[T^0];$$

i.e., according to Section 1.3.2.3.9.3,

$$\mathscr{F}[T]_\xi = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \mathscr{F}[T^0](\boldsymbol{\mu}) \times \delta_{(\boldsymbol{\mu})}.$$

The right-hand side is a *weighted* lattice distribution, whose nodes $\boldsymbol{\mu} \in \mathbb{Z}^n$ are weighted by the *sample values* $\mathscr{F}[T^0](\boldsymbol{\mu})$ of the transform of the motif T^0 at those nodes. Since $T^0 \in \mathscr{E}'$, the latter values may be written

$$\mathscr{F}[T^0](\boldsymbol{\mu}) = \langle T^0_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\mu}\cdot\mathbf{x})\rangle.$$

By the structure theorem for distributions with compact support (Section 1.3.2.3.9.7), T^0 is a derivative of finite order of a continuous function; therefore, from Section 1.3.2.4.2.8 and Section 1.3.2.5.8, $\mathscr{F}[T^0](\boldsymbol{\mu})$ grows at most polynomially as $\|\boldsymbol{\mu}\| \to \infty$ (see also Section 1.3.2.6.10.3 about this property). Conversely, let $W = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}}\delta_{(\boldsymbol{\mu})}$ be a weighted lattice distribution such that the weights $w_{\boldsymbol{\mu}}$ grow at most polynomially as $\|\boldsymbol{\mu}\| \to \infty$. Then W is a tempered distribution, whose Fourier cotransform $T_{\mathbf{x}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \exp(+2\pi i\boldsymbol{\mu}\cdot\mathbf{x})$ is periodic. If T is now written as $r * T^0$ for some $T^0 \in \mathscr{E}'$, then by the reciprocity theorem

$$w_{\boldsymbol{\mu}} = \mathscr{F}[T^0](\boldsymbol{\mu}) = \langle T^0_{\mathbf{x}}, \exp(-2\pi i\boldsymbol{\mu}\cdot\mathbf{x})\rangle.$$

Although the choice of T^0 is not unique, and need not yield back the same motif as may have been used to build T initially, different choices of T^0 will lead to the same coefficients $w_{\boldsymbol{\mu}}$ because of the periodicity of $\exp(-2\pi i\boldsymbol{\mu}\cdot\mathbf{x})$.

The Fourier transformation thus establishes a duality between periodic distributions and weighted lattice distributions. The pair of relations

(i) $$w_{\boldsymbol{\mu}} = \langle T^0_{\mathbf{x}}, \exp(-2\pi i\boldsymbol{\mu}\cdot\mathbf{x})\rangle$$

(ii) $$T_{\mathbf{x}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \exp(+2\pi i\boldsymbol{\mu}\cdot\mathbf{x})$$

are referred to as the *Fourier analysis* and the *Fourier synthesis* of T, respectively (there is a discrepancy between this terminology and the crystallographic one, see Section 1.3.4.2.1.1). In other words, any periodic distribution $T \in \mathscr{S}'$ may be represented by a Fourier series (ii), whose coefficients are calculated by (i). The convergence of (ii) towards T in $\mathscr{S}'$ will be investigated later (Section 1.3.2.6.10).

1.3.2.6.5. The case of nonstandard period lattices

Let Λ denote the nonstandard lattice consisting of all vectors of the form $\sum_{j=1} m_j\mathbf{a}_j$, where the m_j are rational integers and $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are n linearly independent vectors in $\mathbb{R}^n$. Let R be the corresponding lattice distribution: $R = \sum_{x\in\Lambda}\delta_{(\mathbf{x})}$.

Let $\mathbf{A}$ be the nonsingular $n \times n$ matrix whose successive columns are the coordinates of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ in the standard basis of $\mathbb{R}^n$; $\mathbf{A}$ will be called the *period matrix* of Λ, and the mapping $\mathbf{x} \longmapsto \mathbf{Ax}$ will be denoted by A. According to Section 1.3.2.3.9.5 we have

$$\langle R, \varphi\rangle = \sum_{\mathbf{m}\in\mathbb{Z}^n} \varphi(\mathbf{Am}) = \langle r, (A^{-1})^{\#}\varphi\rangle = |\det\mathbf{A}|^{-1}\langle A^{\#}r, \varphi\rangle$$

for any $\varphi \in \mathscr{S}$, and hence $R = |\det\mathbf{A}|^{-1}A^{\#}r$. By Fourier transformation, according to Section 1.3.2.5.5,

$$\mathscr{F}[R] = |\det\mathbf{A}|^{-1}\mathscr{F}[A^{\#}r] = [(\mathbf{A}^{-1})^T]^{\#}\mathscr{F}[r] = [(\mathbf{A}^{-1})^T]^{\#}r,$$

which we write:

$$\mathscr{F}[R] = |\det\mathbf{A}|^{-1}R^*$$

with

$$R^* = |\det\mathbf{A}|[(\mathbf{A}^{-1})^T]^{\#}r.$$

R^* is a lattice distribution:

$$R^* = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n}\delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]} = \sum_{\boldsymbol{\xi}\in\Lambda^*}\delta_{(\boldsymbol{\xi})}$$

associated with the *reciprocal lattice* Λ^* whose basis vectors $\mathbf{a}_1^*, \ldots, \mathbf{a}_n^*$ are the columns of $(\mathbf{A}^{-1})^T$. Since the latter matrix is equal to the adjoint matrix (*i.e.* the matrix of co-factors) of $\mathbf{A}$ divided by $\det\mathbf{A}$, the components of the reciprocal basis vectors can be written down explicitly (see Section 1.3.4.2.1.1 for the crystallographic case $n = 3$).

A distribution T will be called Λ-*periodic* if $\tau_{\boldsymbol{\xi}}T = T$ for all $\boldsymbol{\xi} \in \Lambda$; as previously, T may be written $R * T^0$ for some motif distribution T^0 with compact support. By Fourier transformation,

$$\begin{aligned}\mathscr{F}[T] &= |\det\mathbf{A}|^{-1}R^*\cdot\mathscr{F}[T^0]\\ &= |\det\mathbf{A}|^{-1}\sum_{\boldsymbol{\xi}\in\Lambda^*}\mathscr{F}[T^0](\boldsymbol{\xi})\delta_{(\boldsymbol{\xi})}\\ &= |\det\mathbf{A}|^{-1}\sum_{\boldsymbol{\mu}\in\mathbb{Z}^n}\mathscr{F}[T^0][(\mathbf{A}^{-1})^T\boldsymbol{\mu}]\delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]}\end{aligned}$$

so that $\mathscr{F}[T]$ is a weighted reciprocal-lattice distribution, the weight attached to node $\boldsymbol{\xi} \in \Lambda^*$ being $|\det\mathbf{A}|^{-1}$ times the value $\mathscr{F}[T^0](\boldsymbol{\xi})$ of the Fourier transform of the motif T^0.

This result may be further simplified if T and its motif T^0 are referred to the standard period lattice $\mathbb{Z}^n$ by defining t and t^0 so that $T = A^{\#}t$, $T^0 = A^{\#}t^0$, $t = r * t^0$. Then

$$\mathscr{F}[T^0](\boldsymbol{\xi}) = |\det\mathbf{A}|\mathscr{F}[t^0](\mathbf{A}^T\boldsymbol{\xi}),$$

hence

$$\mathscr{F}[T^0][(\mathbf{A}^{-1})^T\boldsymbol{\mu}] = |\det\mathbf{A}|\mathscr{F}[t^0](\boldsymbol{\mu}),$$

so that

$$\mathscr{F}[T] = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n}\mathscr{F}[t^0](\boldsymbol{\mu})\delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]}$$

in nonstandard coordinates, while

$$\mathscr{F}[t] = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n}\mathscr{F}[t^0](\boldsymbol{\mu})\delta_{(\boldsymbol{\mu})}$$

in standard coordinates.

The reciprocity theorem may then be written:

(iii) $$W_{\boldsymbol{\xi}} = |\det\mathbf{A}|^{-1}\langle T^0_{\mathbf{x}}, \exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\rangle, \quad \boldsymbol{\xi} \in \boldsymbol{\Lambda}^*$$

(iv) $$T_{\mathbf{x}} = \sum_{\boldsymbol{\xi}\in\Lambda^*} W_{\boldsymbol{\xi}} \exp(+2\pi i\boldsymbol{\xi}\cdot\mathbf{x})$$

in nonstandard coordinates, or equivalently:

(v) $$w_{\boldsymbol{\mu}} = \langle t^0_{\mathbf{x}}, \exp(-2\pi i\boldsymbol{\mu}\cdot\mathbf{x})\rangle, \quad \boldsymbol{\mu} \in \mathbb{Z}^n$$

(vi) $$t_{\mathbf{x}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \exp(+2\pi i\boldsymbol{\mu}\cdot\mathbf{x})$$

in standard coordinates. It gives an n-dimensional Fourier series representation for any periodic distribution over $\mathbb{R}^n$. The convergence of such series in $\mathscr{S}'(\mathbb{R}^n)$ will be examined in Section 1.3.2.6.10.

1.3.2.6.6. Duality between periodization and sampling

Let T^0 be a distribution with compact support (the 'motif'). Its Fourier transform $\bar{\mathscr{F}}[T^0]$ is analytic (Section 1.3.2.5.4) and may thus be used as a multiplier.

We may rephrase the preceding results as follows:

(i) if T^0 is 'periodized by R' to give $R * T^0$, then $\bar{\mathscr{F}}[T^0]$ is 'sampled by R^*' to give $|\det \mathbf{A}|^{-1} R^* \cdot \bar{\mathscr{F}}[T^0]$;

(ii) if $\bar{\mathscr{F}}[T^0]$ is 'sampled by R^*' to give $R^* \cdot \bar{\mathscr{F}}[T^0]$, then T^0 is 'periodized by R' to give $|\det \mathbf{A}| R * T^0$.

Thus the Fourier transformation establishes a duality between the periodization of a distribution by a period lattice Λ and the sampling of its transform at the nodes of lattice Λ^* reciprocal to Λ. This is a particular instance of the convolution theorem of Section 1.3.2.5.8.

At this point it is traditional to break the symmetry between $\mathscr{F}$ and $\bar{\mathscr{F}}$ which distribution theory has enabled us to preserve even in the presence of periodicity, and to perform two distinct identifications:

(i) a Λ-periodic distribution T will be handled as a distribution $\tilde{T}$ on $\mathbb{R}^n/\Lambda$, was done in Section 1.3.2.6.3;

(ii) a weighted lattice distribution $W = \sum_{\boldsymbol{\mu} \in \mathbb{Z}^n} W_{\boldsymbol{\mu}} \delta_{[(\mathbf{A}^{-1})^T \boldsymbol{\mu}]}$ will be identified with the collection $\{W_{\boldsymbol{\mu}} | \boldsymbol{\mu} \in \mathbb{Z}^n\}$ of its n-tuply indexed coefficients.

1.3.2.6.7. The Poisson summation formula

Let $\varphi \in \mathscr{S}$, so that $\mathscr{F}[\varphi] \in \mathscr{S}$. Let R be the lattice distribution associated to lattice Λ, with period matrix $\mathbf{A}$, and let R^* be associated to the reciprocal lattice Λ^*. Then we may write:

$$\begin{aligned} \langle R, \varphi \rangle &= \langle R, \bar{\mathscr{F}}[\mathscr{F}[\varphi]] \rangle \\ &= \langle \bar{\mathscr{F}}[R], \mathscr{F}[\varphi] \rangle \\ &= |\det \mathbf{A}|^{-1} \langle R^*, \mathscr{F}[\varphi] \rangle \end{aligned}$$

i.e.

$$\sum_{\mathbf{x} \in \Lambda} \varphi(\mathbf{x}) = |\det \mathbf{A}|^{-1} \sum_{\boldsymbol{\xi} \in \Lambda^*} \mathscr{F}[\varphi](\boldsymbol{\xi}).$$

This identity, which also holds for $\bar{\mathscr{F}}$, is called the *Poisson summation formula*. Its usefulness follows from the fact that the speed of decrease at infinity of φ and $\mathscr{F}[\varphi]$ are inversely related (Section 1.3.2.4.4.3), so that if one of the series (say, the left-hand side) is slowly convergent, the other (say, the right-hand side) will be rapidly convergent. This procedure has been used by Ewald (1921) [see also Bertaut (1952), Born & Huang (1954)] to evaluate lattice sums (Madelung constants) involved in the calculation of the internal electrostatic energy of crystals (see Chapter 3.4 in this volume on convergence acceleration techniques for crystallographic lattice sums and Chapter 3.5 on modern extensions of the Ewald summation method).

When φ is a multivariate Gaussian

$$\varphi(\mathbf{x}) = G_{\mathbf{B}}(\mathbf{x}) = \exp(-\tfrac{1}{2} \mathbf{x}^T \mathbf{B} \mathbf{x}),$$

then

$$\mathscr{F}[\varphi](\boldsymbol{\xi}) = |\det(2\pi \mathbf{B}^{-1})|^{1/2} G_{\mathbf{B}^{-1}}(\boldsymbol{\xi}),$$

and Poisson's summation formula for a lattice with period matrix $\mathbf{A}$ reads:

$$\begin{aligned} \sum_{\mathbf{m} \in \mathbb{Z}^n} G_{\mathbf{B}}(\mathbf{A}\mathbf{m}) = |\det \mathbf{A}|^{-1} |\det(2\pi \mathbf{B}^{-1})|^{1/2} \\ \times \sum_{\boldsymbol{\mu} \in \mathbb{Z}^n} G_{4\pi^2 \mathbf{B}^{-1}}[(\mathbf{A}^{-1})^T \boldsymbol{\mu}] \end{aligned}$$

or equivalently

$$\sum_{\mathbf{m} \in \mathbb{Z}^n} G_{\mathbf{C}}(\mathbf{m}) = |\det(2\pi \mathbf{C}^{-1})|^{1/2} \sum_{\boldsymbol{\mu} \in \mathbb{Z}^n} G_{4\pi^2 \mathbf{C}^{-1}}(\boldsymbol{\mu})$$

with $\mathbf{C} = \mathbf{A}^T \mathbf{B} \mathbf{A}$.

1.3.2.6.8. Convolution of Fourier series

Let $S = R * S^0$ and $T = R * T^0$ be two Λ-periodic distributions, the motifs S^0 and T^0 having compact support. The convolution $S * T$ does not exist, because S and T do not satisfy the support condition (Section 1.3.2.3.9.7). However, the three distributions R, S^0 and T^0 do satisfy the generalized support condition, so that their convolution is defined; then, by associativity and commutativity:

$$R * S^0 * T^0 = S * T^0 = S^0 * T.$$

By Fourier transformation and by the convolution theorem:

$$\begin{aligned} R^* \times \mathscr{F}[S^0 * T^0] &= (R^* \times \mathscr{F}[S^0]) \times \mathscr{F}[T^0] \\ &= \mathscr{F}[T^0] \times (R^* \times \mathscr{F}[S^0]). \end{aligned}$$

Let $\{U_\xi\}_{\xi \in \Lambda^*}$, $\{V_\xi\}_{\xi \in \Lambda^*}$ and $\{W_\xi\}_{\xi \in \Lambda^*}$ be the sets of Fourier coefficients associated to S, T and $S * T^0 (= S^0 * T)$, respectively. Identifying the coefficients of δ_ξ for $\boldsymbol{\xi} \in \Lambda^*$ yields the forward version of the convolution theorem for Fourier series:

$$W_\xi = |\det \mathbf{A}| U_\xi V_\xi.$$

The backward version of the theorem requires that T be infinitely differentiable. The distribution $S \times T$ is then well defined and its Fourier coefficients $\{Q_\xi\}_{\xi \in \Lambda^*}$ are given by

$$Q_\xi = \sum_{\boldsymbol{\eta} \in \Lambda^*} U_{\boldsymbol{\eta}} V_{\xi - \boldsymbol{\eta}}.$$

1.3.2.6.9. Toeplitz forms, Szegö's theorem

Toeplitz forms were first investigated by Toeplitz (1907, 1910, 1911*a*). They occur in connection with the 'trigonometric moment problem' (Shohat & Tamarkin, 1943; Akhiezer, 1965) and probability theory (Grenander, 1952) and play an important role in several direct approaches to the crystallographic phase problem [see Sections 1.3.4.2.1.10, 1.3.4.5.2.2(*e*)]. Many aspects of their theory and applications are presented in the book by Grenander & Szegö (1958).

1.3.2.6.9.1. Toeplitz forms

Let $f \in L^1(\mathbb{R}/\mathbb{Z})$ be real-valued, so that its Fourier coefficients satisfy the relations $c_{-m}(f) = \overline{c_m(f)}$. The Hermitian form in $n+1$ complex variables

$$T_n[f](\mathbf{u}) = \sum_{\mu=0}^{n}\sum_{\nu=0}^{n} \overline{u_\mu} c_{\mu-\nu} u_\nu$$

is called the nth *Toeplitz form* associated to f. It is a straightforward consequence of the convolution theorem and of Parseval's identity that $T_n[f]$ may be written:

$$T_n[f](\mathbf{u}) = \int_0^1 \left| \sum_{\nu=0}^{n} u_\nu \exp(2\pi i \nu x) \right|^2 f(x)\,\mathrm{d}x.$$

1.3.2.6.9.2. The Toeplitz–Carathéodory–Herglotz theorem

It was shown independently by Toeplitz (1911*b*), Carathéodory (1911) and Herglotz (1911) that a function $f \in L^1$ is almost everywhere non-negative *if and only if* the Toeplitz forms $T_n[f]$ associated to f are positive semidefinite for all values of n.

This is equivalent to the infinite system of determinantal inequalities

$$D_n = \det \begin{pmatrix} c_0 & c_{-1} & \cdot & \cdot & c_{-n} \\ c_1 & c_0 & c_{-1} & \cdot & \cdot \\ \cdot & c_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & c_{-1} \\ c_n & \cdot & \cdot & c_1 & c_0 \end{pmatrix} \geq 0 \quad \text{for all } n.$$

The D_n are called *Toeplitz determinants*. Their application to the crystallographic phase problem is described in Section 1.3.4.2.1.10.

1.3.2.6.9.3. Asymptotic distribution of eigenvalues of Toeplitz forms

The eigenvalues of the Hermitian form $T_n[f]$ are defined as the $n+1$ real roots of the characteristic equation $\det\{T_n[f-\lambda]\} = 0$. They will be denoted by

$$\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_{n+1}^{(n)}.$$

It is easily shown that if $m \leq f(x) \leq M$ for all x, then $m \leq \lambda_\nu^{(n)} \leq M$ for all n and all $\nu = 1, \ldots, n+1$. As $n \to \infty$ these bounds, and the distribution of the $\lambda^{(n)}$ within these bounds, can be made more precise by introducing two new notions.

(i) *Essential bounds*: define ess inf f as the largest m such that $f(x) \geq m$ except for values of x forming a set of measure 0; and define ess sup f similarly.

(ii) *Equal distribution*. For each n, consider two sets of $n+1$ real numbers:

$$a_1^{(n)}, a_2^{(n)}, \ldots, a_{n+1}^{(n)}, \quad \text{and} \quad b_1^{(n)}, b_2^{(n)}, \ldots, b_{n+1}^{(n)}.$$

Assume that for each ν and each n, $|a_\nu^{(n)}| < K$ and $|b_\nu^{(n)}| < K$ with K independent of ν and n. The sets $\{a_\nu^{(n)}\}$ and $\{b_\nu^{(n)}\}$ are said to be equally distributed in $[-K, +K]$ if, for any function F over $[-K, +K]$,

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=1}^{n+1} [F(a_\nu^{(n)}) - F(b_\nu^{(n)})] = 0.$$

We may now state an important theorem of Szegö (1915, 1920). Let $f \in L^1$, and put $m = \text{ess inf}\, f$, $M = \text{ess sup}\, f$. If m and M are finite, then for any continuous function $F(\lambda)$ defined in the interval $[m, M]$ we have

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=1}^{n+1} F(\lambda_\nu^{(n)}) = \int_0^1 F[f(x)]\,\mathrm{d}x.$$

In other words, the eigenvalues $\lambda_\nu^{(n)}$ of the T_n and the values $f[\nu/(n+2)]$ of f on a regular subdivision of]0, 1[are equally distributed.

Further investigations into the spectra of Toeplitz matrices may be found in papers by Hartman & Wintner (1950, 1954), Kac *et al.* (1953), Widom (1965), and in the notes by Hirschman & Hughes (1977).

1.3.2.6.9.4. Consequences of Szegö's theorem

(i) If the λ's are ordered in ascending order, then

$$\lim_{n\to\infty} \lambda_1^{(n)} = m = \text{ess inf}\, f, \quad \lim_{n\to\infty} \lambda_{n+1}^{(n)} = M = \text{ess sup}\, f.$$

Thus, when $f \geq 0$, the condition number $\lambda_{n+1}^{(n)}/\lambda_1^{(n)}$ of $T_n[f]$ tends towards the 'essential dynamic range' M/m of f.

(ii) Let $F(\lambda) = \lambda^s$ where s is a positive integer. Then

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=1}^{n+1} [\lambda_\nu^{(n)}]^s = \int_0^1 [f(x)]^s\,\mathrm{d}x.$$

(iii) Let $m > 0$, so that $\lambda_\nu^{(n)} > 0$, and let $D_n(f) = \det T_n(f)$. Then

$$D_n(f) = \prod_{\nu=1}^{n+1} \lambda_\nu^{(n)},$$

hence

$$\log D_n(f) = \sum_{\nu=1}^{n+1} \log \lambda_\nu^{(n)}.$$

Putting $F(\lambda) = \log \lambda$, it follows that

$$\lim_{n\to\infty} [D_n(f)]^{1/(n+1)} = \exp\left\{ \int_0^1 \log f(x)\,\mathrm{d}x \right\}.$$

Further terms in this limit were obtained by Szegö (1952) and interpreted in probabilistic terms by Kac (1954).

1.3.2.6.10. Convergence of Fourier series

The investigation of the convergence of Fourier series and of more general trigonometric series has been the subject of intense study for over 150 years [see *e.g.* Zygmund (1976)]. It has been a constant source of new mathematical ideas and theories, being directly responsible for the birth of such fields as set theory, topology and functional analysis.

This section will briefly survey those aspects of the classical results in dimension 1 which are relevant to the practical use of Fourier series in crystallography. The books by Zygmund (1959), Tolstov (1962) and Katznelson (1968) are standard references in the field, and Dym & McKean (1972) is recommended as a stimulant.

1.3.2.6.10.1. Classical L^1 theory

The space $L^1(\mathbb{R}/\mathbb{Z})$ consists of (equivalence classes of) complex-valued functions f on the circle which are summable, *i.e.* for which

$$\|f\|_1 \equiv \int_0^1 |f(x)|\,\mathrm{d}x < +\infty.$$

It is a convolution algebra: If f and g are in L^1, then $f * g$ is in L^1.

The mth Fourier coefficient $c_m(f)$ of f,

$$c_m(f) = \int_0^1 f(x)\exp(-2\pi imx)\,\mathrm{d}x$$

is bounded: $|c_m(f)| \le \|f\|_1$, and by the Riemann–Lebesgue lemma $c_m(f) \to 0$ as $m \to \infty$. By the convolution theorem, $c_m(f * g) = c_m(f)c_m(g)$.

The pth *partial sum* $S_p(f)$ of the Fourier series of f,

$$S_p(f)(x) = \sum_{|m|\le p} c_m(f)\exp(2\pi imx),$$

may be written, by virtue of the convolution theorem, as $S_p(f) = D_p * f$, where

$$D_p(x) = \sum_{|m|\le p}\exp(2\pi imx) = \frac{\sin[(2p+1)\pi x]}{\sin \pi x}$$

is the *Dirichlet kernel*. Because D_p comprises numerous slowly decaying oscillations, both positive and negative, $S_p(f)$ may not converge towards f in a strong sense as $p \to \infty$. Indeed, spectacular pathologies are known to exist where the partial sums, examined pointwise, diverge everywhere (Zygmund, 1959, Chapter VIII). When f is piecewise continuous, but presents isolated jumps, convergence near these jumps is marred by the *Gibbs phenomenon*: $S_p(f)$ always 'overshoots the mark' by about 9%, the area under the spurious peak tending to 0 as $p \to \infty$ but not its height [see Larmor (1934) for the history of this phenomenon].

By contrast, the *arithmetic mean* of the partial sums, also called the pth Cesàro sum,

$$C_p(f) = \frac{1}{p+1}[S_0(f) + \ldots + S_p(f)],$$

converges to f in the sense of the L^1 norm: $\|C_p(f) - f\|_1 \to 0$ as $p \to \infty$. If furthermore f is *continuous*, then the convergence is *uniform*, *i.e.* the error is bounded everywhere by a quantity which goes to 0 as $p \to \infty$. It may be shown that

$$C_p(f) = F_p * f,$$

where

$$F_p(x) = \sum_{|m|\le p}\left(1 - \frac{|m|}{p+1}\right)\exp(2\pi imx) = \frac{1}{p+1}\left[\frac{\sin(p+1)\pi x}{\sin \pi x}\right]^2$$

is the *Fejér kernel*. F_p has over D_p the advantage of being everywhere positive, so that the Cesàro sums $C_p(f)$ of a positive function f are always positive.

The de la Vallée Poussin kernel

$$V_p(x) = 2F_{2p+1}(x) - F_p(x)$$

has a trapezoidal distribution of coefficients and is such that $c_m(V_p) = 1$ if $|m| \le p+1$; therefore $V_p * f$ is a trigonometric polynomial with the same Fourier coefficients as f over that range of values of m.

The Poisson kernel

$$P_r(x) = 1 + 2\sum_{m=1}^{\infty} r^m \cos 2\pi mx = \frac{1 - r^2}{1 - 2r\cos 2\pi mx + r^2}$$

with $0 \le r < 1$ gives rise to an Abel summation procedure [Tolstov (1962, p. 162); Whittaker & Watson (1927, p. 57)] since

$$(P_r * f)(x) = \sum_{m\in\mathbb{Z}} c_m(f) r^{|m|}\exp(2\pi imx).$$

Compared with the other kernels, P_r has the disadvantage of not being a trigonometric polynomial; however, P_r is the real part of the Cauchy kernel (Cartan, 1961; Ahlfors, 1966):

$$P_r(x) = \mathscr{R}e\left[\frac{1 + r\exp(2\pi ix)}{1 - r\exp(2\pi ix)}\right]$$

and hence provides a link between trigonometric series and analytic functions of a complex variable.

Other methods of summation involve forming a moving average of f by convolution with other sequences of functions $\alpha_p(\mathbf{x})$ besides D_p of F_p which 'tend towards δ' as $p \to \infty$. The convolution is performed by multiplying the Fourier coefficients of f by those of α_p, so that one forms the quantities

$$S'_p(f)(x) = \sum_{|m|\le p} c_m(\alpha_p)c_m(f)\exp(2\pi imx).$$

For instance the 'sigma factors' of Lanczos (Lanczos, 1966, p. 65), defined by

$$\sigma_m = \frac{\sin[m\pi/p]}{m\pi/p},$$

lead to a summation procedure whose behaviour is intermediate between those using the Dirichlet and the Fejér kernels; it corresponds to forming a moving average of f by convolution with

$$\alpha_p = p\,\chi_{[-1/(2p),\,1/(2p)]} * D_p,$$

which is itself the convolution of a 'rectangular pulse' of width $1/p$ and of the Dirichlet kernel of order p.

A review of the summation problem in crystallography is given in Section 1.3.4.2.1.3.

1.3.2.6.10.2. Classical L^2 theory

The space $L^2(\mathbb{R}/\mathbb{Z})$ of (equivalence classes of) square-integrable complex-valued functions f on the circle is contained in $L^1(\mathbb{R}/\mathbb{Z})$, since by the Cauchy–Schwarz inequality

$$\|f\|_1^2 = \left(\int_0^1 |f(x)| \times 1 \, \mathrm{d}x\right)^2 \leq \left(\int_0^1 |f(x)|^2 \, \mathrm{d}x\right)\left(\int_0^1 1^2 \, \mathrm{d}x\right) = \|f\|_2^2 \leq \infty.$$

Thus all the results derived for L^1 hold for L^2, a great simplification over the situation in $\mathbb{R}$ or $\mathbb{R}^n$ where neither L^1 nor L^2 was contained in the other.

However, more can be proved in L^2, because L^2 is a Hilbert space (Section 1.3.2.2.4) for the inner product

$$(f, g) = \int_0^1 \overline{f(x)} g(x) \, \mathrm{d}x,$$

and because the family of functions $\{\exp(2\pi imx)\}_{m\in\mathbb{Z}}$ constitutes an orthonormal Hilbert basis for L^2.

The sequence of Fourier coefficients $c_m(f)$ of $f \in L^2$ belongs to the space $\ell^2(\mathbb{Z})$ of square-summable sequences:

$$\sum_{m\in\mathbb{Z}} |c_m(f)|^2 < \infty.$$

Conversely, every element $c = (c_m)$ of ℓ^2 is the sequence of Fourier coefficients of a unique function in L^2. The inner product

$$(c, d) = \sum_{m\in\mathbb{Z}} \overline{c_m} d_m$$

makes ℓ^2 into a Hilbert space, and the map from L^2 to ℓ^2 established by the Fourier transformation is an isometry (Parseval/Plancherel):

$$\|f\|_{L^2} = \|c(f)\|_{\ell^2}$$

or equivalently:

$$(f, g) = (c(f), c(g)).$$

This is a useful property in applications, since (f, g) may be calculated either from f and g themselves, or from their Fourier coefficients $c(f)$ and $c(g)$ (see Section 1.3.4.4.6) for crystallographic applications).

By virtue of the orthogonality of the basis $\{\exp(2\pi imx)\}_{m\in\mathbb{Z}}$, the partial sum $S_p(f)$ is the best mean-square fit to f in the linear subspace of L^2 spanned by $\{\exp(2\pi imx)\}_{|m|\leq p}$, and hence (Bessel's inequality)

$$\sum_{|m|\leq p} |c_m(f)|^2 = \|f\|_2^2 - \sum_{|M|\geq p} |c_M(f)|^2 \leq \|f\|_2^2.$$

1.3.2.6.10.3. The viewpoint of distribution theory

The use of distributions enlarges considerably the range of behaviour which can be accommodated in a Fourier series, even in the case of general dimension n where classical theories meet with even more difficulties than in dimension 1.

Let $\{w_m\}_{m\in\mathbb{Z}}$ be a sequence of complex numbers with $|w_m|$ growing at most polynomially as $|m| \to \infty$, say $|w_m| \leq C|m|^K$. Then the sequence $\{w_m/(2\pi im)^{K+2}\}_{m\in\mathbb{Z}}$ is in ℓ^2 and even defines a *continuous* function $f \in L^2(\mathbb{R}/\mathbb{Z})$ and an associated tempered distribution $T_f \in \mathscr{D}'(\mathbb{R}/\mathbb{Z})$. Differentiation of T_f $(K + 2)$ times then yields a tempered distribution whose Fourier transform leads to the original sequence of coefficients. Conversely, by the structure theorem for distributions with compact support (Section 1.3.2.3.9.7), the motif T^0 of a $\mathbb{Z}$-periodic distribution is a derivative of finite order of a continuous function; hence its Fourier coefficients will grow at most polynomially with $|m|$ as $|m| \to \infty$.

Thus distribution theory allows the manipulation of Fourier series whose coefficients exhibit polynomial growth as their order goes to infinity, while those derived from functions had to tend to 0 by virtue of the Riemann–Lebesgue lemma. The distribution-theoretic approach to Fourier series holds even in the case of general dimension n, where classical theories meet with even more difficulties (see Ash, 1976) than in dimension 1.

1.3.2.7. The discrete Fourier transformation

1.3.2.7.1. Shannon's sampling theorem and interpolation formula

Let $\varphi \in \mathscr{E}(\mathbb{R}^n)$ be such that $\Phi = \mathscr{F}[\varphi]$ has compact support K. Let φ be sampled at the nodes of a lattice Λ^*, yielding the lattice distribution $R^* \times \varphi$. The Fourier transform of this sampled version of φ is

$$\mathscr{F}[R^* \times \varphi] = |\det \mathbf{A}|(R * \Phi),$$

which is essentially Φ periodized by period lattice $\Lambda = (\Lambda^*)^*$, with period matrix $\mathbf{A}$.

Let us assume that Λ is such that the translates of K by different period vectors of Λ are disjoint. Then we may recover Φ from $R * \Phi$ by masking the contents of a 'unit cell' $\mathscr{V}$ of Λ (*i.e.* a fundamental domain for the action of Λ in $\mathbb{R}^n$) whose boundary does not meet K. If $\chi_{\mathscr{V}}$ is the indicator function of $\mathscr{V}$, then

$$\Phi = \chi_{\mathscr{V}} \times (R * \Phi).$$

Transforming both sides by $\bar{\mathscr{F}}$ yields

$$\varphi = \bar{\mathscr{F}}\left[\chi_{\mathscr{V}} \times \frac{1}{|\det \mathbf{A}|} \mathscr{F}[R^* \times \varphi]\right],$$

i.e.

$$\varphi = \left(\frac{1}{V} \bar{\mathscr{F}}[\chi_{\mathscr{V}}]\right) * (R^* \times \varphi)$$

since $|\det \mathbf{A}|$ is the volume V of $\mathscr{V}$.

This interpolation formula is traditionally credited to Shannon (1949), although it was discovered much earlier by Whittaker (1915). It shows that φ may be recovered from its sample values on Λ^* (*i.e.* from $R^* \times \varphi$) provided Λ^* is sufficiently fine that no overlap (or 'aliasing') occurs in the periodization of Φ by the dual lattice Λ. The interpolation kernel is the transform of the normalized indicator function of a unit cell of Λ containing the support K of Φ.

If K is contained in a sphere of radius $1/\Delta$ and if Λ and Λ^* are rectangular, the length of each basis vector of Λ must be greater than $2/\Delta$, and thus the sampling interval must be smaller than $\Delta/2$. This requirement constitutes the Shannon sampling criterion.

1.3.2.7.2. Duality between subdivision and decimation of period lattices

1.3.2.7.2.1. Geometric description of sublattices

Let $\Lambda_{\mathbf{A}}$ be a period lattice in $\mathbb{R}^n$ with matrix $\mathbf{A}$, and let $\Lambda^*_{\mathbf{A}}$ be the lattice reciprocal to $\Lambda_{\mathbf{A}}$, with period matrix $(A^{-1})^T$. Let $\Lambda_{\mathbf{B}}$, $\mathbf{B}$, $\Lambda^*_{\mathbf{B}}$ be defined similarly, and let us suppose that $\Lambda_{\mathbf{A}}$ *is a sublattice of* $\Lambda_{\mathbf{B}}$, *i.e.* that $\Lambda_{\mathbf{B}} \supset \Lambda_{\mathbf{A}}$ as a set.

The relation between $\Lambda_{\mathbf{A}}$ and $\Lambda_{\mathbf{B}}$ may be described in two different fashions: (i) multiplicatively, and (ii) additively.

(i) We may write $\mathbf{A} = \mathbf{BN}$ for some nonsingular matrix $\mathbf{N}$ with integer entries. $\mathbf{N}$ may be viewed as the period matrix of the coarser lattice $\Lambda_{\mathbf{A}}$ with respect to the period basis of the finer lattice $\Lambda_{\mathbf{B}}$. It will be more convenient to write $\mathbf{A} = \mathbf{DB}$, where $\mathbf{D} = \mathbf{BNB}^{-1}$ is a rational matrix (with integer determinant since $\det \mathbf{D} = \det \mathbf{N}$) in terms of which the two lattices are related by

$$\Lambda_{\mathbf{A}} = \mathbf{D}\Lambda_{\mathbf{B}}.$$

(ii) Call two vectors in $\Lambda_{\mathbf{B}}$ *congruent modulo* $\Lambda_{\mathbf{A}}$ if their difference lies in $\Lambda_{\mathbf{A}}$. Denote the set of congruence classes (or 'cosets') by $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$, and the number of these classes by $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$. The 'coset decomposition'

$$\Lambda_{\mathbf{B}} = \bigcup_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} (\ell + \Lambda_{\mathbf{A}})$$

represents $\Lambda_{\mathbf{B}}$ as the disjoint union of $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$ translates of $\Lambda_{\mathbf{A}}$. $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ is a finite lattice with $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$ elements, called the *residual lattice* of $\Lambda_{\mathbf{B}}$ modulo $\Lambda_{\mathbf{A}}$.

The two descriptions are connected by the relation $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}] = \det \mathbf{D} = \det \mathbf{N}$, which follows from a volume calculation. We may also combine (i) and (ii) into

$$\text{(iii)} \qquad \Lambda_{\mathbf{B}} = \bigcup_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} (\ell + \mathbf{D}\Lambda_{\mathbf{B}})$$

which may be viewed as the n-dimensional equivalent of the Euclidean algorithm for integer division: ℓ is the 'remainder' of the division by $\Lambda_{\mathbf{A}}$ of a vector in $\Lambda_{\mathbf{B}}$, the quotient being the matrix $\mathbf{D}$.

1.3.2.7.2.2. Sublattice relations for reciprocal lattices

Let us now consider the two *reciprocal lattices* $\Lambda^*_{\mathbf{A}}$ and $\Lambda^*_{\mathbf{B}}$. Their period matrices $(\mathbf{A}^{-1})^T$ and $(\mathbf{B}^{-1})^T$ are related by: $(\mathbf{B}^{-1})^T = (\mathbf{A}^{-1})^T \mathbf{N}^T$, where $\mathbf{N}^T$ is an integer matrix; or equivalently by $(\mathbf{B}^{-1})^T = \mathbf{D}^T (\mathbf{A}^{-1})^T$. This shows that the roles are reversed in that $\Lambda^*_{\mathbf{B}}$ *is a sublattice of* $\Lambda^*_{\mathbf{A}}$, which we may write:

$$\text{(i)}^* \qquad \Lambda^*_{\mathbf{B}} = \mathbf{D}^T \Lambda^*_{\mathbf{A}}$$

$$\text{(ii)}^* \qquad \Lambda^*_{\mathbf{A}} = \bigcup_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} (\ell^* + \Lambda^*_{\mathbf{B}}).$$

The residual lattice $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$ is finite, with $[\Lambda^*_{\mathbf{A}} : \Lambda^*_{\mathbf{B}}] = \det \mathbf{D} = \det \mathbf{N} = [\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$, and we may again combine (i)* and (ii)* into

$$\text{(iii)}^* \qquad \Lambda^*_{\mathbf{A}} = \bigcup_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} (\ell^* + \mathbf{D}^T \Lambda^*_{\mathbf{A}}).$$

1.3.2.7.2.3. Relation between lattice distributions

The above relations between lattices may be rewritten in terms of the corresponding *lattice distributions* as follows:

$$\text{(i)} \qquad R_{\mathbf{A}} = \frac{1}{|\det \mathbf{D}|} \mathbf{D}^{\#} R^*_{\mathbf{B}}$$

$$\text{(ii)} \qquad R_{\mathbf{B}} = T_{\mathbf{B/A}} * R_{\mathbf{A}}$$

$$\text{(i)}^* \qquad R^*_{\mathbf{B}} = \frac{1}{|\det \mathbf{D}|} (\mathbf{D}^T)^{\#} R^*_{\mathbf{A}}$$

$$\text{(ii)}^* \qquad R^*_{\mathbf{A}} = T^*_{\mathbf{A/B}} * R^*_{\mathbf{B}}$$

where

$$T_{\mathbf{B/A}} = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \delta_{(\ell)}$$

and

$$T^*_{\mathbf{A/B}} = \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \delta_{(\ell^*)}$$

are (finite) residual-lattice distributions. We may incorporate the factor $1/|\det \mathbf{D}|$ in (i) and (i)* into these distributions and define

$$S_{\mathbf{B/A}} = \frac{1}{|\det \mathbf{D}|} T_{\mathbf{B/A}}, \quad S^*_{\mathbf{A/B}} = \frac{1}{|\det \mathbf{D}|} T^*_{\mathbf{A/B}}.$$

Since $|\det \mathbf{D}| = [\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}] = [\Lambda^*_{\mathbf{A}} : \Lambda^*_{\mathbf{B}}]$, convolution with $S_{\mathbf{B/A}}$ and $S^*_{\mathbf{A/B}}$ has the effect of *averaging* the translates of a distribution under the elements (or 'cosets') of the residual lattices $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ and $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, respectively. This process will be called 'coset averaging'. Eliminating $R_{\mathbf{A}}$ and $R_{\mathbf{B}}$ between (i) and (ii), and $R^*_{\mathbf{A}}$ and $R^*_{\mathbf{B}}$ between (i)* and (ii)*, we may write:

$$\text{(i}') \qquad R_{\mathbf{A}} = \mathbf{D}^{\#}(S_{\mathbf{B/A}} * R_{\mathbf{A}})$$

$$\text{(ii}') \qquad R_{\mathbf{B}} = S_{\mathbf{B/A}} * (\mathbf{D}^{\#} R_{\mathbf{B}})$$

$$\text{(i}')^* \qquad R^*_{\mathbf{B}} = (\mathbf{D}^T)^{\#}(S^*_{\mathbf{A/B}} * R^*_{\mathbf{B}})$$

$$\text{(ii}')^* \qquad R^*_{\mathbf{A}} = S^*_{\mathbf{A/B}} * [(\mathbf{D}^T)^{\#} R^*_{\mathbf{A}}].$$

These identities show that *period subdivision* by convolution with $S_{\mathbf{B/A}}$ (respectively $S^*_{\mathbf{A/B}}$) on the one hand, and *period decimation* by 'dilation' by $\mathbf{D}^{\#}$ on the other hand, are mutually inverse operations on $R_{\mathbf{A}}$ and $R_{\mathbf{B}}$ (respectively $R^*_{\mathbf{A}}$ and $R^*_{\mathbf{B}}$).

1.3.2.7.2.4. Relation between Fourier transforms

Finally, let us consider the relations between the *Fourier transforms* of these lattice distributions. Recalling the basic relation of Section 1.3.2.6.5,

$$\begin{aligned} \mathscr{F}[R_{\mathbf{A}}] &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{A}} \\ &= \frac{1}{|\det \mathbf{DB}|} T^*_{\mathbf{A/B}} * R^*_{\mathbf{B}} \qquad \text{by (ii)}^* \\ &= \left(\frac{1}{|\det \mathbf{D}|} T^*_{\mathbf{A/B}} \right) * \left(\frac{1}{|\det \mathbf{B}|} R^*_{\mathbf{B}} \right) \end{aligned}$$

i.e.

$$\text{(iv)} \qquad \mathscr{F}[R_{\mathbf{A}}] = S^*_{\mathbf{A/B}} * \mathscr{F}[R_{\mathbf{B}}]$$

and similarly:

$$\text{(v)} \qquad \mathscr{F}[R^*_{\mathbf{B}}] = S_{\mathbf{B/A}} * \mathscr{F}[R^*_{\mathbf{A}}].$$

Thus $R_{\mathbf{A}}$ (respectively $R^*_{\mathbf{B}}$), a *decimated* version of $R_{\mathbf{B}}$ (respectively $R^*_{\mathbf{A}}$), is transformed by $\mathscr{F}$ into a *subdivided* version of $\mathscr{F}[R_{\mathbf{B}}]$ (respectively $\mathscr{F}[R^*_{\mathbf{A}}]$).

The converse is also true:

$$\begin{aligned}\mathscr{F}[R_{\mathbf{B}}] &= \frac{1}{|\det \mathbf{B}|} R^*_{\mathbf{B}} \\ &= \frac{1}{|\det \mathbf{B}|}\frac{1}{|\det \mathbf{D}|}(\mathbf{D}^T)^{\#} R^*_{\mathbf{A}} \qquad \text{by (i)}^* \\ &= (\mathbf{D}^T)^{\#}\left(\frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{A}}\right)\end{aligned}$$

i.e.

$$\text{(iv}') \qquad \mathscr{F}[R_{\mathbf{B}}] = (\mathbf{D}^T)^{\#}\mathscr{F}[R_{\mathbf{A}}]$$

and similarly

$$\text{(v}') \qquad \mathscr{F}[R^*_{\mathbf{A}}] = \mathbf{D}^{\#}\mathscr{F}[R^*_{\mathbf{B}}].$$

Thus $R_{\mathbf{B}}$ (respectively $R^*_{\mathbf{A}}$), a *subdivided* version of $R_{\mathbf{A}}$ (respectively $R^*_{\mathbf{B}}$) is transformed by $\mathscr{F}$ into a *decimated* version of $\mathscr{F}[R_{\mathbf{A}}]$ (respectively $\mathscr{F}[R^*_{\mathbf{B}}]$). Therefore, *the Fourier transform exchanges subdivision and decimation of period lattices for lattice distributions*.

Further insight into this phenomenon is provided by applying $\bar{\mathscr{F}}$ to both sides of (iv) and (v) and invoking the convolution theorem:

$$\text{(iv}'') \qquad R_{\mathbf{A}} = \bar{\mathscr{F}}[S^*_{\mathbf{A/B}}] \times R_{\mathbf{B}}$$

$$\text{(v}'') \qquad R^*_{\mathbf{B}} = \bar{\mathscr{F}}[S_{\mathbf{B/A}}] \times R^*_{\mathbf{A}}.$$

These identities show that multiplication by the transform of the period-subdividing distribution $S^*_{\mathbf{A/B}}$ (respectively $S_{\mathbf{B/A}}$) has the effect of decimating $R_{\mathbf{B}}$ to $R_{\mathbf{A}}$ (respectively $R^*_{\mathbf{A}}$ to $R^*_{\mathbf{B}}$). They clearly imply that, if $\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ and $\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, then

$$\begin{aligned}\bar{\mathscr{F}}[S^*_{\mathbf{A/B}}](\ell) &= 1 \text{ if } \ell = \mathbf{0} \quad (\textit{i.e.} \text{ if } \ell \text{ belongs} \\ &\qquad\qquad \text{to the class of } \Lambda_{\mathbf{A}}), \\ &= 0 \text{ if } \ell \neq \mathbf{0}; \\ \bar{\mathscr{F}}[S_{\mathbf{B/A}}](\ell^*) &= 1 \text{ if } \ell^* = \mathbf{0} \quad (\textit{i.e.} \text{ if } \ell^* \text{ belongs} \\ &\qquad\qquad \text{to the class of } \Lambda^*_{\mathbf{B}}), \\ &= 0 \text{ if } \ell^* \neq \mathbf{0}.\end{aligned}$$

Therefore, the duality between subdivision and decimation may be viewed as another aspect of that between convolution and multiplication.

There is clearly a strong analogy between the sampling/periodization duality of Section 1.3.2.6.6 and the decimation/subdivision duality, which is viewed most naturally in terms of subgroup relationships: both sampling and decimation involve restricting a function to a *discrete additive subgroup* of the domain over which it is initially given.

1.3.2.7.2.5. Sublattice relations in terms of periodic distributions

The usual presentation of this duality is not in terms of lattice distributions, but of periodic distributions obtained by convolving them with a motif.

Given $T^0 \in \mathscr{E}'(\mathbb{R}^n)$, let us form $R_{\mathbf{A}} * T^0$, then *decimate* its transform $(1/|\det \mathbf{A}|)R^*_{\mathbf{A}} \times \bar{\mathscr{F}}[T^0]$ by keeping only its values at the points of the coarser lattice $\Lambda^*_{\mathbf{B}} = \mathbf{D}^T\Lambda^*_{\mathbf{A}}$; as a result, $R^*_{\mathbf{A}}$ is replaced by $(1/|\det \mathbf{D}|)R^*_{\mathbf{B}}$, and the reverse transform then yields

$$\frac{1}{|\det \mathbf{D}|} R_{\mathbf{B}} * T^0 = S_{\mathbf{B/A}} * (R_{\mathbf{A}} * T^0) \qquad \text{by (ii)},$$

which is the *coset-averaged* version of the original $R_{\mathbf{A}} * T^0$. The converse situation is analogous to that of Shannon's sampling theorem. Let a function $\varphi \in \mathscr{E}(\mathbb{R}^n)$ whose transform $\Phi = \mathscr{F}[\varphi]$ has compact support be sampled as $R_{\mathbf{B}} \times \varphi$ at the nodes of $\Lambda_{\mathbf{B}}$. Then

$$\mathscr{F}[R_{\mathbf{B}} \times \varphi] = \frac{1}{|\det \mathbf{B}|}(R^*_{\mathbf{B}} * \Phi)$$

is periodic with period lattice $\Lambda^*_{\mathbf{B}}$. If the sampling lattice $\Lambda_{\mathbf{B}}$ is decimated to $\Lambda_{\mathbf{A}} = \mathbf{D}\Lambda_{\mathbf{B}}$, the inverse transform becomes

$$\begin{aligned}\mathscr{F}[R_{\mathbf{A}} \times \varphi] &= \frac{1}{|\det \mathbf{D}|}(R^*_{\mathbf{A}} * \Phi) \\ &= S^*_{\mathbf{A/B}} * (R^*_{\mathbf{B}} * \Phi) \qquad \text{by (ii)}^*,\end{aligned}$$

hence becomes periodized more finely by averaging over the cosets of $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$. With this finer periodization, the various copies of Supp Φ may start to overlap (a phenomenon called 'aliasing'), indicating that decimation has produced too coarse a sampling of φ.

1.3.2.7.3. Discretization of the Fourier transformation

Let $\varphi^0 \in \mathscr{E}(\mathbb{R}^n)$ be such that $\Phi^0 = \mathscr{F}[\varphi^0]$ has compact support (φ^0 is said to be *band-limited*). Then $\varphi = R_{\mathbf{A}} * \varphi^0$ is $\Lambda_{\mathbf{A}}$-periodic, and $\Phi = \mathscr{F}[\varphi] = (1/|\det \mathbf{A}|)R^*_{\mathbf{A}} \times \Phi^0$ is such that only a finite number of points $\lambda^*_{\mathbf{A}}$ of $\Lambda^*_{\mathbf{A}}$ have a nonzero Fourier coefficient $\Phi^0(\lambda^*_{\mathbf{A}})$ attached to them. We may therefore find a *decimation* $\Lambda^*_{\mathbf{B}} = \mathbf{D}^T\Lambda^*_{\mathbf{A}}$ of $\Lambda^*_{\mathbf{A}}$ such that the distinct translates of Supp Φ^0 by vectors of $\Lambda^*_{\mathbf{B}}$ do not intersect.

The distribution Φ can be uniquely recovered from $R^*_{\mathbf{B}} * \Phi$ by the procedure of Section 1.3.2.7.1, and we may write:

$$\begin{aligned}R^*_{\mathbf{B}} * \Phi &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{B}} * (R^*_{\mathbf{A}} \times \Phi^0) \\ &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{A}} \times (R^*_{\mathbf{B}} * \Phi^0) \\ &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{B}} * [T^*_{\mathbf{A/B}} \times (R^*_{\mathbf{B}} * \Phi^0)];\end{aligned}$$

these rearrangements being legitimate because Φ^0 and $T^*_{\mathbf{A/B}}$ have compact supports which are intersection-free under the action of $\Lambda^*_{\mathbf{B}}$. By virtue of its $\Lambda^*_{\mathbf{B}}$-periodicity, this distribution is entirely characterized by its 'motif' $\tilde{\Phi}$ with respect to $\Lambda^*_{\mathbf{B}}$:

$$\tilde{\Phi} = \frac{1}{|\det \mathbf{A}|} T^*_{\mathbf{A/B}} \times (R^*_{\mathbf{B}} * \Phi^0).$$

Similarly, φ may be uniquely recovered by Shannon interpolation from the distribution sampling its values at the nodes of $\Lambda_{\mathbf{B}} = \mathbf{D}^{-1}\Lambda_{\mathbf{A}}$($\Lambda_{\mathbf{B}}$ is a *subdivision* of $\Lambda_{\mathbf{B}}$). By virtue of its $\Lambda_{\mathbf{A}}$-periodicity, this distribution is completely characterized by its motif:

$$\tilde{\varphi} = T_{\mathbf{B/A}} \times \varphi = T_{\mathbf{B/A}} \times (R^*_{\mathbf{A}} * \varphi^0).$$

Let $\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ and $\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, and define the two sets of coefficients

(1) $\tilde{\varphi}(\ell) = \varphi(\ell + \boldsymbol{\lambda}_{\mathbf{A}})$ for any $\boldsymbol{\lambda}_{\mathbf{A}} \in \Lambda_{\mathbf{A}}$ (all choices of $\boldsymbol{\lambda}_{\mathbf{A}}$ give the same $\tilde{\varphi}$),

(2) $\tilde{\Phi}(\ell^*) = \Phi^0(\ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}})$ for the unique $\boldsymbol{\lambda}^*_{\mathbf{B}}$ (if it exists) such that $\ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}} \in \text{Supp } \Phi^0$,
$= 0$ if no such $\boldsymbol{\lambda}^*_{\mathbf{B}}$ exists.

Define the two distributions

$$\omega = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \tilde{\varphi}(\ell)\delta_{(\ell)}$$

and

$$\Omega = \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \tilde{\Phi}(\ell^*)\delta_{(\ell^*)}.$$

The relation between ω and Ω has two equivalent forms:

(i) $R_{\mathbf{A}} * \omega = \mathscr{F}[R^*_{\mathbf{B}} * \Omega]$

(ii) $\bar{\mathscr{F}}[R_{\mathbf{A}} * \omega] = R^*_{\mathbf{B}} * \Omega.$

By (i), $R_{\mathbf{A}} * \omega = |\det \mathbf{B}| R_{\mathbf{B}} \times \mathscr{F}[\Omega]$. Both sides are weighted lattice distributions concentrated at the nodes of $\Lambda_{\mathbf{B}}$, and equating the weights at $\boldsymbol{\lambda}_{\mathbf{B}} = \ell + \boldsymbol{\lambda}_{\mathbf{A}}$ gives

$$\tilde{\varphi}(\ell) = \frac{1}{|\det \mathbf{D}|} \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \tilde{\Phi}(\ell^*) \exp[-2\pi i \ell^* \cdot (\ell + \boldsymbol{\lambda}_{\mathbf{A}})].$$

Since $\ell^* \in \Lambda^*_{\mathbf{A}}$, $\ell^* \cdot \boldsymbol{\lambda}_{\mathbf{A}}$ is an integer, hence

$$\tilde{\varphi}(\ell) = \frac{1}{|\det \mathbf{D}|} \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \tilde{\Phi}(\ell^*) \exp(-2\pi i \ell^* \cdot \ell).$$

By (ii), we have

$$\frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{B}} * [T^*_{\mathbf{A/B}} \times (R^*_{\mathbf{B}} * \Phi^0)] = \frac{1}{|\det \mathbf{A}|} \bar{\mathscr{F}}[R_{\mathbf{A}} * \omega].$$

Both sides are weighted lattice distributions concentrated at the nodes of $\Lambda^*_{\mathbf{B}}$, and equating the weights at $\boldsymbol{\lambda}^*_{\mathbf{A}} = \ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}}$ gives

$$\tilde{\Phi}(\ell^*) = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \tilde{\varphi}(\ell) \exp[+2\pi i \ell \cdot (\ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}})].$$

Since $\ell \in \Lambda_{\mathbf{B}}$, $\ell \cdot \boldsymbol{\lambda}^*_{\mathbf{B}}$ is an integer, hence

$$\tilde{\Phi}(\ell^*) = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \tilde{\varphi}(\ell) \exp(+2\pi i \ell \cdot \ell^*).$$

Now the decimation/subdivision relations between $\Lambda_{\mathbf{A}}$ and $\Lambda_{\mathbf{B}}$ may be written:

$$\mathbf{A} = \mathbf{DB} = \mathbf{BN},$$

so that

$$\ell = \mathbf{B}\mathscr{k} \quad \text{for } \mathscr{k} \in \mathbb{Z}^n$$
$$\ell^* = (\mathbf{A}^{-1})^T \mathscr{k}^* \quad \text{for } \mathscr{k}^* \in \mathbb{Z}^n$$

with $(\mathbf{A}^{-1})^T = (\mathbf{B}^{-1})^T(\mathbf{N}^{-1})^T$, hence finally

$$\ell^* \cdot \ell = \ell \cdot \ell^* = \mathscr{k}^* \cdot (\mathbf{N}^{-1}\mathscr{k}).$$

Denoting $\tilde{\varphi}(\mathbf{B}\mathscr{k})$ by $\psi(\mathscr{k})$ and $\tilde{\Phi}[(\mathbf{A}^{-1})^T\mathscr{k}^*]$ by $\Psi(\mathscr{k}^*)$, the relation between ω and Ω may be written in the equivalent form

(i) $$\psi(\mathscr{k}) = \frac{1}{|\det \mathbf{N}|} \sum_{\mathscr{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n} \Psi(\mathscr{k}^*) \exp[-2\pi i \mathscr{k}^* \cdot (\mathbf{N}^{-1}\mathscr{k})]$$

(ii) $$\Psi(\mathscr{k}^*) = \sum_{\mathscr{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n} \psi(\mathscr{k}) \exp[+2\pi i \mathscr{k}^* \cdot (\mathbf{N}^{-1}\mathscr{k})],$$

where the summations are now over *finite* residual lattices in standard form.

Equations (i) and (ii) describe two mutually inverse linear transformations $\mathscr{F}(\mathbf{N})$ and $\bar{\mathscr{F}}(\mathbf{N})$ between two vector spaces $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$ of dimension $|\det \mathbf{N}|$. $\mathscr{F}(\mathbf{N})$ [respectively $\bar{\mathscr{F}}(\mathbf{N})$] is the *discrete* Fourier (respectively inverse Fourier) transform associated to matrix $\mathbf{N}$.

The vector spaces $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$ may be viewed from two different standpoints:

(1) as vector spaces of *weighted residual-lattice distributions*, of the form $\alpha(\mathbf{x})T_{\mathbf{B/A}}$ and $\beta(\mathbf{x})T^*_{\mathbf{A/B}}$; the canonical basis of $W_{\mathbf{N}}$ (respectively $W^*_{\mathbf{N}}$) then consists of the $\delta_{(\mathscr{k})}$ for $\mathscr{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ [respectively $\delta_{(\mathscr{k}^*)}$ for $\mathscr{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$];

(2) as vector spaces of *weight vectors* for the $|\det \mathbf{N}|$ δ-functions involved in the expression for $T_{\mathbf{B/A}}$ (respectively $T^*_{\mathbf{A/B}}$); the canonical basis of $W_{\mathbf{N}}$ (respectively $W^*_{\mathbf{N}}$) consists of weight vectors $\mathbf{u}_{\mathscr{k}}$ (respectively $\mathbf{v}_{\mathscr{k}^*}$) giving weight 1 to element $\mathscr{k}$ (respectively $\mathscr{k}^*$) and 0 to the others.

These two spaces are said to be 'isomorphic' (a relation denoted $\cong$), the isomorphism being given by the one-to-one correspondence:

$$\omega = \sum_{\mathscr{k}} \psi(\mathscr{k})\delta_{(\mathscr{k})} \quad \leftrightarrow \quad \psi = \sum_{\mathscr{k}} \psi(\mathscr{k})\mathbf{u}_{\mathscr{k}}$$
$$\Omega = \sum_{\mathscr{k}^*} \Psi(\mathscr{k}^*)\delta_{(\mathscr{k}^*)} \quad \leftrightarrow \quad \Psi = \sum_{\mathscr{k}^*} \Psi(\mathscr{k}^*)\mathbf{v}_{\mathscr{k}^*}.$$

The second viewpoint will be adopted, as it involves only linear algebra. However, it is most helpful to keep the first one in mind and to think of the data or results of a discrete Fourier transform as representing (through their sets of unique weights) two periodic lattice distributions related by the full, distribution-theoretic Fourier transform.

We therefore view $W_{\mathbf{N}}$ (respectively $W^*_{\mathbf{N}}$) as the vector space of complex-valued functions over the finite residual lattice $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ (respectively $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$) and write:

$$W_{\mathbf{N}} \cong L(\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}) \cong L(\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n)$$
$$W_{\mathbf{N}}^* \cong L(\Lambda_{\mathbf{A}}^*/\Lambda_{\mathbf{B}}^*) \cong L(\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n)$$

since a vector such as ψ is in fact the function $\mathbf{k} \longmapsto \psi(\mathbf{k})$.

The two spaces $W_{\mathbf{N}}$ and $W_{\mathbf{N}}^*$ may be equipped with the following Hermitian inner products:

$$(\varphi, \psi)_W = \sum_{\mathbf{k}} \overline{\varphi(\mathbf{k})}\psi(\mathbf{k})$$
$$(\Phi, \Psi)_{W^*} = \sum_{\mathbf{k}} \overline{\Phi(\mathbf{k}^*)}\Psi(\mathbf{k}^*),$$

which makes each of them into a *Hilbert space*. The canonical bases $\{\mathbf{u}_{\mathbf{k}} | \mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n\}$ and $\{\mathbf{v}_{\mathbf{k}^*} | \mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n\}$ and $W_{\mathbf{N}}$ and $W_{\mathbf{N}}^*$ are *orthonormal* for their respective product.

1.3.2.7.4. Matrix representation of the discrete Fourier transform (DFT)

By virtue of definitions (i) and (ii),

$$\mathscr{F}(\mathbf{N})\mathbf{v}_{\mathbf{k}^*} = \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{k}} \exp[-2\pi i \mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]\mathbf{u}_{\mathbf{k}}$$
$$\bar{\mathscr{F}}(\mathbf{N})\mathbf{u}_{\mathbf{k}} = \sum_{\mathbf{k}^*} \exp[+2\pi i \mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]\mathbf{v}_{\mathbf{k}^*}$$

so that $\mathscr{F}(\mathbf{N})$ and $\bar{\mathscr{F}}(\mathbf{N})$ may be represented, in the canonical bases of $W_{\mathbf{N}}$ and $W_{\mathbf{N}}^*$, by the following matrices:

$$[\mathscr{F}(\mathbf{N})]_{\mathbf{k}\mathbf{k}^*} = \frac{1}{|\det \mathbf{N}|} \exp[-2\pi i \mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]$$
$$[\bar{\mathscr{F}}(\mathbf{N})]_{\mathbf{k}^*\mathbf{k}} = \exp[+2\pi i \mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})].$$

When $\mathbf{N}$ is symmetric, $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ and $\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ may be identified in a natural manner, and the above matrices are symmetric.

When $\mathbf{N}$ is diagonal, say $\mathbf{N} = \mathrm{diag}(\nu_1, \nu_2, \ldots, \nu_n)$, then the tensor product structure of the full multidimensional Fourier transform (Section 1.3.2.4.2.4)

$$\mathscr{F}_{\mathbf{x}} = \mathscr{F}_{x_1} \otimes \mathscr{F}_{x_2} \otimes \ldots \otimes \mathscr{F}_{x_n}$$

gives rise to a tensor product structure for the DFT matrices. The tensor product of matrices is defined as follows:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \ldots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{n1}\mathbf{B} & \ldots & a_{nn}\mathbf{B} \end{pmatrix}.$$

Let the index vectors $\mathbf{k}$ and $\mathbf{k}^*$ be ordered in the same way as the elements in a Fortran array, *e.g.* for $\mathbf{k}$ with k_1 increasing fastest, k_2 next fastest, ..., k_n slowest; then

$$\mathscr{F}(\mathbf{N}) = \mathscr{F}(\nu_1) \otimes \mathscr{F}(\nu_2) \otimes \ldots \otimes \mathscr{F}(\nu_n),$$

where

$$[\mathscr{F}(\nu_j)]_{k_j, k_j^*} = \frac{1}{\nu_j} \exp\left(-2\pi i \frac{k_j^* k_j}{\nu_j}\right),$$

and

$$\bar{\mathscr{F}}(\mathbf{N}) = \bar{\mathscr{F}}(\nu_1) \otimes \bar{\mathscr{F}}(\nu_2) \otimes \ldots \otimes \bar{\mathscr{F}}(\nu_n),$$

where

$$[\bar{\mathscr{F}}_{\nu_j}]_{k_j^*, k_j} = \exp\left(+2\pi i \frac{k_j^* k_j}{\nu_j}\right).$$

1.3.2.7.5. Properties of the discrete Fourier transform

The DFT inherits most of the properties of the Fourier transforms, but with certain numerical factors ('Jacobians') due to the transition from continuous to discrete measure.

(1) *Linearity* is obvious.

(2) *Shift property*. If $(\tau_a \psi)(\mathbf{k}) = \psi(\mathbf{k} - \mathbf{a})$ and $(\tau_{a^*}\Psi)(\mathbf{k}^*) = \Psi(\mathbf{k}^* - \mathbf{a}^*)$, where subtraction takes place by modular vector arithmetic in $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ and $\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$, respectively, then the following identities hold:

$$\bar{\mathscr{F}}(\mathbf{N})[\tau_{\mathbf{k}}\psi](\mathbf{k}^*) = \exp[+2\pi i \mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]\bar{\mathscr{F}}(\mathbf{N})[\psi](\mathbf{k}^*)$$
$$\mathscr{F}(\mathbf{N})[\tau_{\mathbf{k}^*}\Psi](\mathbf{k}) = \exp[-2\pi i \mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]\mathscr{F}(\mathbf{N})[\Psi](\mathbf{k}).$$

(3) *Differentiation identities*. Let vectors $\boldsymbol{\psi}$ and $\boldsymbol{\Psi}$ be constructed from $\varphi^0 \in \mathscr{E}(\mathbb{R}^n)$ as in Section 1.3.2.7.3, hence be related by the DFT. If $D^{\mathbf{p}}\boldsymbol{\psi}$ designates the vector of sample values of $D_{\mathbf{x}}^{\mathbf{p}}\varphi^0$ at the points of $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$, and $D^{\mathbf{p}}\boldsymbol{\Psi}$ the vector of values of $D_{\boldsymbol{\xi}}^{\mathbf{p}}\Phi^0$ at points of $\Lambda_{\mathbf{A}}^*/\Lambda_{\mathbf{B}}^*$, then for all multi-indices $\mathbf{p} = (p_1, p_2, \ldots, p_n)$

$$(D^{\mathbf{p}}\boldsymbol{\psi})(\mathbf{k}) = \bar{\mathscr{F}}(\mathbf{N})[(+2\pi i \mathbf{k}^*)^{\mathbf{p}}\boldsymbol{\Psi}](\mathbf{k})$$
$$(D^{\mathbf{p}}\boldsymbol{\Psi})(\mathbf{k}^*) = \mathscr{F}(\mathbf{N})[(-2\pi i \mathbf{k})^{\mathbf{p}}\boldsymbol{\psi}](\mathbf{k}^*)$$

or equivalently

$$\mathscr{F}(\mathbf{N})[D^{\mathbf{p}}\boldsymbol{\psi}](\mathbf{k}^*) = (+2\pi i \mathbf{k}^*)^{\mathbf{p}}\boldsymbol{\Psi}(\mathbf{k}^*)$$
$$\bar{\mathscr{F}}(\mathbf{N})[D^{\mathbf{p}}\boldsymbol{\Psi}](\mathbf{k}) = (-2\pi i \mathbf{k})^{\mathbf{p}}\boldsymbol{\psi}(\mathbf{k}).$$

(4) *Convolution property*. Let $\boldsymbol{\varphi} \in W_{\mathbf{N}}$ and $\boldsymbol{\Phi} \in W_{\mathbf{N}}^*$ (respectively $\boldsymbol{\psi}$ and $\boldsymbol{\Psi}$) be related by the DFT, and define

$$(\boldsymbol{\varphi} * \boldsymbol{\psi})(\mathbf{k}) = \sum_{\mathbf{k}' \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n} \boldsymbol{\varphi}(\mathbf{k}')\boldsymbol{\psi}(\mathbf{k} - \mathbf{k}')$$
$$(\boldsymbol{\Phi} * \boldsymbol{\Psi})(\mathbf{k}^*) = \sum_{\mathbf{k}^{*'} \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n} \boldsymbol{\Phi}(\mathbf{k}^{*'})\boldsymbol{\Psi}(\mathbf{k}^* - \mathbf{k}^{*'}).$$

Then

$$\bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\Phi} * \boldsymbol{\Psi}](\mathbf{k}) = |\det \mathbf{N}|\boldsymbol{\varphi}(\mathbf{k})\boldsymbol{\psi}(\mathbf{k})$$
$$\mathscr{F}(\mathbf{N})[\boldsymbol{\varphi} * \boldsymbol{\psi}](\mathbf{k}^*) = \boldsymbol{\Phi}(\mathbf{k}^*)\boldsymbol{\Psi}(\mathbf{k}^*)$$

and

$$\bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\varphi} \times \boldsymbol{\psi}](\mathbf{k}^*) = \frac{1}{|\det \mathbf{N}|}(\boldsymbol{\Phi} * \boldsymbol{\Psi})(\mathbf{k}^*)$$
$$\mathscr{F}(\mathbf{N})[\boldsymbol{\Phi} \times \boldsymbol{\Psi}](\mathbf{k}) = (\boldsymbol{\varphi} * \boldsymbol{\psi})(\mathbf{k}).$$

Since addition on $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ and $\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ is modular, this type of convolution is called *cyclic* convolution.

(5) *Parseval/Plancherel property*. If $\boldsymbol{\varphi}$, $\boldsymbol{\psi}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ are as above, then

$$(\mathscr{F}(\mathbf{N})[\boldsymbol{\Phi}], \mathscr{F}(\mathbf{N})[\boldsymbol{\Psi}])_W = \frac{1}{|\det \mathbf{N}|}(\boldsymbol{\Phi}, \boldsymbol{\Psi})_{W^*}$$
$$(\bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\varphi}], \bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\psi}])_W = \frac{1}{|\det \mathbf{N}|}(\boldsymbol{\varphi}, \boldsymbol{\psi})_W.$$

(6) *Period* 4. When $\mathbf{N}$ is symmetric, so that the ranges of indices $\mathscr{k}$ and $\mathscr{k}^*$ can be identified, it makes sense to speak of powers of $\mathscr{F}(\mathbf{N})$ and $\bar{\mathscr{F}}(\mathbf{N})$. Then the 'standardized' matrices $(1/|\det \mathbf{N}|^{1/2})\mathscr{F}(\mathbf{N})$ and $(1/|\det \mathbf{N}|^{1/2})\bar{\mathscr{F}}(\mathbf{N})$ are *unitary* matrices whose fourth power is the identity matrix (Section 1.3.2.4.3.4); their eigenvalues are therefore ± 1 and $\pm i$.

1.3.3. Numerical computation of the discrete Fourier transform

1.3.3.1. Introduction

The Fourier transformation's most remarkable property is undoubtedly that of turning convolution into multiplication. As distribution theory has shown, other valuable properties – such as the shift property, the conversion of differentiation into multiplication by monomials, and the duality between periodicity and sampling – are special instances of the convolution theorem.

This property is exploited in many areas of applied mathematics and engineering (Campbell & Foster, 1948; Sneddon, 1951; Champeney, 1973; Bracewell, 1986). For example, the passing of a signal through a linear filter, which results in its being convolved with the response of the filter to a δ-function 'impulse', may be modelled as a multiplication of the signal's transform by the transform of the impulse response (also called transfer function). Similarly, the solution of systems of partial differential equations may be turned by Fourier transformation into a division problem for distributions. In both cases, the formulations obtained after Fourier transformation are considerably simpler than the initial ones, and lend themselves to constructive solution techniques.

Whenever the functions to which the Fourier transform is applied are band-limited, or can be well approximated by band-limited functions, the discrete Fourier transform (DFT) provides a means of constructing explicit numerical solutions to the problems at hand. A great variety of investigations in physics, engineering and applied mathematics thus lead to DFT calculations, to such a degree that, at the time of writing, about 50% of all supercomputer CPU time is alleged to be spent calculating DFTs.

The straightforward use of the defining formulae for the DFT leads to calculations of size N^2 for N sample points, which become unfeasible for any but the smallest problems. Much ingenuity has therefore been exerted on the design and implementation of faster algorithms for calculating the DFT (McClellan & Rader, 1979; Nussbaumer, 1981; Blahut, 1985; Brigham, 1988). The most famous is that of Cooley & Tukey (1965) which heralded the age of digital signal processing. However, it had been preceded by the prime factor algorithm of Good (1958, 1960), which has lately been the basis of many new developments. Recent historical research (Goldstine, 1977, pp. 249–253; Heideman *et al.*, 1984) has shown that Gauss essentially knew the Cooley–Tukey algorithm as early as 1805 (before Fourier's 1807 work on harmonic analysis!); while it has long been clear that Dirichlet knew of the basis of the prime factor algorithm and used it extensively in his theory of multiplicative characters [see *e.g.* Chapter I of Ayoub (1963), and Chapters 6 and 8 of Apostol (1976)]. Thus the computation of the DFT, far from being a purely technical and rather narrow piece of specialized numerical analysis, turns out to have very rich connections with such central areas of pure mathematics as number theory (algebraic and analytic), the representation theory of certain Lie groups and coding theory – to list only a few. The interested reader may consult Auslander & Tolimieri (1979); Auslander, Feig & Winograd (1982, 1984); Auslander & Tolimieri (1985); Tolimieri (1985).

One-dimensional algorithms are examined first. The Sande mixed-radix version of the Cooley–Tukey algorithm only calls upon the additive structure of congruence classes of integers. The prime factor algorithm of Good begins to exploit some of their multiplicative structure, and the use of relatively prime factors leads to a stronger factorization than that of Sande. Fuller use of the multiplicative structure, *via* the group of units, leads to the Rader algorithm; and the factorization of short convolutions then yields the Winograd algorithms.

Multidimensional algorithms are at first built as tensor products of one-dimensional elements. The problem of factoring the DFT in several dimensions simultaneously is then examined. The section ends with a survey of attempts at formalizing the interplay between algorithm structure and computer architecture for the purpose of automating the design of optimal DFT code.

It was originally intended to incorporate into this section a survey of all the basic notions and results of abstract algebra which are called upon in the course of these developments, but time limitations have made this impossible. This material, however, is adequately covered by the first chapter of Tolimieri *et al.* (1989) in a form tailored for the same purposes. Similarly, the inclusion of numerous detailed examples of the algorithms described here has had to be postponed to a later edition, but an abundant supply of such examples may be found in the signal processing literature, for instance in the books by McClellan & Rader (1979), Blahut (1985), and Tolimieri *et al.* (1989).

1.3.3.2. One-dimensional algorithms

Throughout this section we will denote by $e(t)$ the expression $\exp(2\pi i t)$, $t \in \mathbb{R}$. The mapping $t \longmapsto e(t)$ has the following properties:

$$e(t_1 + t_2) = e(t_1)e(t_2)$$
$$e(-t) = \overline{e(t)} = [e(t)]^{-1}$$
$$e(t) = 1 \Leftrightarrow t \in \mathbb{Z}.$$

Thus e defines an isomorphism between the additive group $\mathbb{R}/\mathbb{Z}$ (the reals modulo the integers) and the multiplicative group of complex numbers of modulus 1. It follows that the mapping $\ell \longmapsto e(\ell/N)$, where $\ell \in \mathbb{Z}$ and N is a positive integer, defines an isomorphism between the one-dimensional residual lattice $\mathbb{Z}/N\mathbb{Z}$ and the multiplicative group of Nth roots of unity.

The DFT on N points then relates vectors $\mathbf{X}$ and $\mathbf{X}^*$ in W and W^* through the linear transformations:

$$F(N): \quad X(k) = \frac{1}{N}\sum_{k^* \in \mathbb{Z}/N\mathbb{Z}} X^*(k^*)e(-k^*k/N)$$
$$\bar{F}(N): \quad X^*(k^*) = \sum_{k \in \mathbb{Z}/N\mathbb{Z}} X(k)e(k^*k/N).$$

1.3.3.2.1. The Cooley–Tukey algorithm

The presentation of Gentleman & Sande (1966) will be followed first [see also Cochran *et al.* (1967)]. It will then be reinterpreted in geometric terms which will prepare the way for the treatment of multidimensional transforms in Section 1.3.3.3.

Suppose that the number of sample points N is composite, say $N = N_1N_2$. We may write k to the base N_1 and k^* to the base N_2 as follows:

$$k = k_1 + N_1 k_2 \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2 \in \mathbb{Z}/N_2\mathbb{Z}$$
$$k^* = k_2^* + k_1^* N_2 \quad k_1^* \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2^* \in \mathbb{Z}/N_2\mathbb{Z}.$$

The defining relation for $\bar{F}(N)$ may then be written:

$$X^*(k_2^* + k_1^* N_2) = \sum_{k_1 \in \mathbb{Z}/N_1\mathbb{Z}} \sum_{k_2 \in \mathbb{Z}/N_2\mathbb{Z}} X(k_1 + N_1 k_2) \times e\left[\frac{(k_2^* + k_1^* N_2)(k_1 + N_1 k_2)}{N_1 N_2}\right].$$

The argument of $e[.]$ may be expanded as

$$\frac{k_2^* k_1}{N} + \frac{k_1^* k_1}{N_1} + \frac{k_2^* k_2}{N_2} + k_1^* k_2,$$

and the last summand, being an integer, may be dropped:

$$X^*(k_2^* + k_1^* N_2) = \sum_{k_1} \left\{ e\left(\frac{k_2^* k_1}{N}\right) \left[\sum_{k_2} X(k_1 + N_1 k_2) e\left(\frac{k_2^* k_2}{N_2}\right) \right] \right\} \times e\left(\frac{k_1^* k_1}{N_1}\right).$$

This computation may be decomposed into five stages, as follows:

(i) form the N_1 vectors $\mathbf{Y}_{k_1}$ of length N_2 by the prescription

$$Y_{k_1}(k_2) = X(k_1 + N_1 k_2), \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2 \in \mathbb{Z}/N_2\mathbb{Z};$$

(ii) calculate the N_1 transforms $\mathbf{Y}_{k_1}^*$ on N_2 points:

$$\mathbf{Y}_{k_1}^* = \bar{F}(N_2)[\mathbf{Y}_{k_1}], \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z};$$

(iii) form the N_2 vectors $\mathbf{Z}_{k_2^*}$ of length N_1 by the prescription

$$\mathbf{Z}_{k_2^*}(k_1) = e\left(\frac{k_2^* k_1}{N}\right) Y_{k_1}^*(k_2^*), \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2^* \in \mathbb{Z}/N_2\mathbb{Z};$$

(iv) calculate the N_2 transforms $\mathbf{Z}_{k_2^*}^*$ on N_1 points:

$$\mathbf{Z}_{k_2^*}^* = \bar{F}(N_1)[\mathbf{Z}_{k_2^*}], \quad k_2^* \in \mathbb{Z}/N_2\mathbb{Z};$$

(v) collect $X^*(k_2^* + k_1^* N_2)$ as $Z_{k_2^*}^*(k_1^*)$.

If the intermediate transforms in stages (ii) and (iv) are performed *in place*, *i.e.* with the results overwriting the data, then at stage (v) the result $X^*(k_2^* + k_1^* N_2)$ will be found at address $k_1^* + N_1 k_2^*$. This phenomenon is called *scrambling* by 'digit reversal', and stage (v) is accordingly known as *unscrambling*.

The initial *N-point* transform $\bar{F}(N)$ has thus been performed as N_1 transforms $\bar{F}(N_2)$ on N_2 points, followed by N_2 transforms $\bar{F}(N_1)$ on N_1 points, thereby reducing the arithmetic cost from $(N_1 N_2)^2$ to $N_1 N_2 (N_1 + N_2)$. The phase shifts applied at stage (iii) are traditionally called 'twiddle factors', and the transposition between k_1 and k_2^* can be performed by the fast recursive technique of Eklundh (1972). Clearly, this procedure can be applied recursively if N_1 and N_2 are themselves composite, leading to an overall arithmetic cost of order $N \log N$ if N has no large prime factors.

The Cooley–Tukey factorization may also be derived from a geometric rather than arithmetic argument. The decomposition $k = k_1 + N_1 k_2$ is associated to a geometric partition of the residual lattice $\mathbb{Z}/N\mathbb{Z}$ into N_1 copies of $\mathbb{Z}/N_2\mathbb{Z}$, each translated by $k_1 \in \mathbb{Z}/N_1\mathbb{Z}$ and 'blown up' by a factor N_1. This partition in turn induces a (direct sum) decomposition of $\mathbf{X}$ as

$$\mathbf{X} = \sum_{k_1} \mathbf{X}_{k_1},$$

where

$$X_{k_1}(k) = X(k) \quad \text{if } k \equiv k_1 \bmod N_1,$$
$$= 0 \qquad \text{otherwise.}$$

According to (i), $\mathbf{X}_{k_1}$ is related to $\mathbf{Y}_{k_1}$ by *decimation by* N_1 and *offset by* k_1. By Section 1.3.2.7.2, $\bar{F}(N)[\mathbf{X}_{k_1}]$ is related to $\bar{F}(N_2)[\mathbf{Y}_{k_1}]$ by *periodization by* N_2 and *phase shift by* $e(k^* k_1/N)$, so that

$$X^*(k^*) = \sum_{k_1} e\left(\frac{k^* k_1}{N}\right) Y_{k_1}^*(k_2^*),$$

the periodization by N_2 being reflected by the fact that $Y_{k_1}^*$ does not depend on k_1^*. Writing $k^* = k_2^* + k_1^* N_2$ and expanding $k^* k_1$ shows that the phase shift contains both the twiddle factor $e(k_2^* k_1/N)$ and the kernel $e(k_1^* k_1/N_1)$ of $\bar{F}(N_1)$. The Cooley–Tukey algorithm is thus naturally associated to the coset decomposition of a lattice modulo a sublattice (Section 1.3.2.7.2).

It is readily seen that essentially the same factorization can be obtained for $F(N)$, up to the complex conjugation of the twiddle factors. The normalizing constant $1/N$ arises from the normalizing constants $1/N_1$ and $1/N_2$ in $F(N_1)$ and $F(N_2)$, respectively.

Factors of 2 are particularly simple to deal with and give rise to a characteristic computational structure called a 'butterfly loop'. If $N = 2M$, then two options exist:

(*a*) using $N_1 = 2$ and $N_2 = M$ leads to collecting the even-numbered coordinates of $\mathbf{X}$ into $\mathbf{Y}_0$ and the odd-numbered coordinates into $\mathbf{Y}_1$

$$Y_0(k_2) = X(2k_2), \qquad k_2 = 0, \ldots, M-1,$$
$$Y_1(k_2) = X(2k_2 + 1), \quad k_2 = 0, \ldots, M-1,$$

and writing:

$$X^*(k_2^*) = Y_0^*(k_2^*) + e(k_2^*/N) Y_1^*(k_2^*), \quad k_2^* = 0, \ldots, M-1;$$
$$X^*(k_2^* + M) = Y_0^*(k_2^*) - e(k_2^*/N) Y_1^*(k_2^*), \quad k_2^* = 0, \ldots, M-1.$$

This is the original version of Cooley & Tukey, and the process of formation of $\mathbf{Y}_0$ and $\mathbf{Y}_1$ is referred to as 'decimation in time' (*i.e.* decimation along the *data* index $\mathbf{k}$).

(*b*) using $N_1 = M$ and $N_2 = 2$ leads to forming

$$Z_0(k_1) = X(k_1) + X(k_1 + M), \qquad k_1 = 0, \ldots, M-1,$$
$$Z_1(k_1) = [X(k_1) - X(k_1 + M)] e\left(\frac{k_1}{N}\right), \qquad k_1 = 0, \ldots, M-1,$$

then obtaining separately the even-numbered and odd-numbered components of $\mathbf{X}^*$ by transforming $\mathbf{Z}_0$ and $\mathbf{Z}_1$:

$$X^*(2k_1^*) = Z_0^*(k_1^*), \quad k_1^* = 0, \ldots, M-1;$$
$$X^*(2k_1^*+1) = Z_1^*(k_1^*), \quad k_1^* = 0, \ldots, M-1.$$

This version is due to Sande (Gentleman & Sande, 1966), and the process of separately obtaining even-numbered and odd-numbered results has led to its being referred to as 'decimation in frequency' (*i.e.* decimation along the *result* index k^*).

By repeated factoring of the number N of sample points, the calculation of $F(N)$ and $\bar{F}(N)$ can be reduced to a succession of stages, the smallest of which operate on single prime factors of N. The reader is referred to Gentleman & Sande (1966) for a particularly lucid analysis of the programming considerations which help implement this factorization efficiently; see also Singleton (1969). Powers of two are often grouped together into factors of 4 or 8, which are advantageous in that they require fewer complex multiplications than the repeated use of factors of 2. In this approach, large prime factors P are detrimental, since they require a full P^2-size computation according to the defining formula.

1.3.3.2.2. The Good (or prime factor) algorithm

1.3.3.2.2.1. Ring structure on $\mathbb{Z}/N\mathbb{Z}$

The set $\mathbb{Z}/N\mathbb{Z}$ of congruence classes of integers modulo an integer N [see *e.g.* Apostol (1976), Chapter 5] inherits from $\mathbb{Z}$ not only the additive structure used in deriving the Cooley–Tukey factorization, but also a *multiplicative* structure in which the product of two congruence classes mod N is uniquely defined as the class of the ordinary product (in $\mathbb{Z}$) of representatives of each class. The multiplication can be distributed over addition in the usual way, endowing $\mathbb{Z}/N\mathbb{Z}$ with the structure of a *commutative ring*.

If N is composite, the ring $\mathbb{Z}/N\mathbb{Z}$ has *zero divisors*. For example, let $N = N_1N_2$, let $n_1 \equiv N_1 \bmod N$, and let $n_2 \equiv N_2 \bmod N$: then $n_1n_2 \equiv 0 \bmod N$. In the general case, a product of nonzero elements will be zero whenever these elements collect together all the factors of N. These circumstances give rise to a fundamental theorem in the theory of commutative rings, the *Chinese Remainder Theorem* (CRT), which will now be stated and proved [see Apostol (1976), Chapter 5; Schroeder (1986), Chapter 16].

1.3.3.2.2.2. The Chinese remainder theorem

Let $N = N_1N_2 \ldots N_d$ be factored into a product of pairwise coprime integers, so that g.c.d. $(N_i, N_j) = 1$ for $i \neq j$. Then the system of congruence equations

$$\ell \equiv \ell_j \bmod N_j, \qquad j = 1, \ldots, d,$$

has a unique solution $\ell \bmod N$. In other words, each $\ell \in \mathbb{Z}/N\mathbb{Z}$ is associated in a one-to-one fashion to the d-tuple $(\ell_1, \ell_2, \ldots, \ell_d)$ of its residue classes in $\mathbb{Z}/N_1\mathbb{Z}, \mathbb{Z}/N_2\mathbb{Z}, \ldots, \mathbb{Z}/N_d\mathbb{Z}$.

The proof of the CRT goes as follows. Let

$$Q_j = \frac{N}{N_j} = \prod_{i \neq j} N_i.$$

Since g.c.d. $(N_j, Q_j) = 1$ there exist integers n_j and q_j such that

$$n_jN_j + q_jQ_j = 1, \qquad j = 1, \ldots, d,$$

then the integer

$$\ell = \sum_{i=1}^{d} \ell_i q_i Q_i \bmod N$$

is the solution. Indeed,

$$\ell \equiv \ell_j q_j Q_j \bmod N_j$$

because all terms with $i \neq j$ contain N_j as a factor; and

$$q_jQ_j \equiv 1 \bmod N_j$$

by the defining relation for q_j.

It may be noted that

$$(q_iQ_i)(q_jQ_j) \equiv 0 \qquad \bmod N \text{ for } i \neq j,$$
$$(q_jQ_j)^2 \equiv q_jQ_j \qquad \bmod N, \ j = 1, \ldots, d,$$

so that the q_jQ_j are mutually orthogonal *idempotents* in the ring $\mathbb{Z}/N\mathbb{Z}$, with properties formally similar to those of mutually orthogonal *projectors onto subspaces* in linear algebra. The analogy is exact, since by virtue of the CRT the ring $\mathbb{Z}/N\mathbb{Z}$ may be considered as the direct product

$$\mathbb{Z}/N_1\mathbb{Z} \times \mathbb{Z}/N_2\mathbb{Z} \times \ldots \times \mathbb{Z}/N_d\mathbb{Z}$$

via the two mutually inverse mappings:

(i) $\ell \longmapsto (\ell_1, \ell_2, \ldots, \ell_d)$ by $\ell \equiv \ell_j \bmod N_j$ for each j;

(ii) $(\ell_1, \ell_2, \ldots, \ell_d) \longmapsto \ell$ by $\ell = \sum_{i=1}^{d} \ell_i q_i Q_i \bmod N$.

The mapping defined by (ii) is sometimes called the 'CRT reconstruction' of ℓ from the ℓ_j.

These two mappings have the property of sending sums to sums and products to products, *i.e*:

(i) $\ell + \ell' \longmapsto (\ell_1 + \ell_1', \ell_2 + \ell_2', \ldots, \ell_d + \ell_d')$
$\ell\ell' \longmapsto (\ell_1\ell_1', \ell_2\ell_2', \ldots, \ell_d\ell_d')$

(ii) $(\ell_1 + \ell_1', \ell_2 + \ell_2', \ldots, \ell_d + \ell_d') \longmapsto \ell + \ell'$
$(\ell_1\ell_1', \ell_2\ell_2', \ldots, \ell_d\ell_d') \longmapsto \ell\ell'$

(the last proof requires using the properties of the idempotents q_jQ_j). This may be described formally by stating that the CRT establishes a *ring isomorphism*:

$$\mathbb{Z}/N\mathbb{Z} \cong (\mathbb{Z}/N_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/N_d\mathbb{Z}).$$

1.3.3.2.2.3. The prime factor algorithm

The CRT will now be used to factor the N-point DFT into a tensor product of d transforms, the jth of length N_j.

Let the indices k and k^* be subjected to the following mappings:

(i) $k \longmapsto (k_1, k_2, \ldots, k_d)$, $k_j \in \mathbb{Z}/N_j\mathbb{Z}$, by $k_j \equiv k \bmod N_j$ for each j, with reconstruction formula

$$k = \sum_{i=1}^{d} k_i q_i Q_i \bmod N;$$

(ii) $k^* \longmapsto (k_1^*, k_2^*, \ldots, k_d^*)$, $k_j^* \in \mathbb{Z}/N_j\mathbb{Z}$, by $k_j^* \equiv q_jk^* \bmod N_j$ for each j, with reconstruction formula

$$k^* = \sum_{i=1}^{d} k_i^* Q_i \bmod N.$$

Then

$$k^*k = \left(\sum_{i=1}^{d} k_i^* Q_i\right)\left(\sum_{j=1}^{d} k_j q_j Q_j\right) \bmod N$$
$$= \sum_{i,j=1}^{d} k_i^* k_j Q_i q_j Q_j \bmod N.$$

Cross terms with $i \neq j$ vanish since they contain all the factors of N, hence

$$k^*k = \sum_{j=1}^{d} q_j Q_j^2 k_j^* k_j \bmod N$$
$$= \sum_{j=1}^{d} (1 - n_j N_j) Q_j k_j^* k_j \bmod N.$$

Dividing by N, which may be written as $N_j Q_j$ for each j, yields

$$\frac{k^*k}{N} = \sum_{j=1}^{d} (1 - n_j N_j) \frac{Q_j}{N_j Q_j} k_j^* k_j \bmod 1$$
$$= \sum_{j=1}^{d} \left(\frac{1}{N_j} - n_j\right) k_j^* k_j \bmod 1,$$

and hence

$$\frac{k^*k}{N} \equiv \sum_{j=1}^{d} \frac{k_j^* k_j}{N_j} \bmod 1.$$

Therefore, by the multiplicative property of $e(.)$,

$$e\left(\frac{k^*k}{N}\right) \equiv \bigotimes_{j=1}^{d} e\left(\frac{k_j^* k_j}{N_j}\right).$$

Let $\mathbf{X} \in L(\mathbb{Z}/N\mathbb{Z})$ be described by a one-dimensional array $X(k)$ indexed by k. The index mapping (i) turns $\mathbf{X}$ into an element of $L(\mathbb{Z}/N_1\mathbb{Z} \times \ldots \times \mathbb{Z}/N_d\mathbb{Z})$ described by a d-dimensional array $X(k_1, \ldots, k_d)$; the latter may be transformed by $\bar{F}(N_1) \bigotimes \ldots \bigotimes \bar{F}(N_d)$ into a new array $X^*(k_1^*, k_2^*, \ldots, k_d^*)$. Finally, the one-dimensional array of results $X^*(k^*)$ will be obtained by reconstructing k^* according to (ii).

The prime factor algorithm, like the Cooley–Tukey algorithm, reindexes a 1D transform to turn it into d separate transforms, but the use of coprime factors and CRT index mapping leads to the further gain that *no twiddle factors* need to be applied between the successive transforms (see Good, 1971). This makes up for the cost of the added complexity of the CRT index mapping.

The natural factorization of N for the prime factor algorithm is thus its factorization into prime powers: $\bar{F}(N)$ is then the tensor product of separate transforms (one for each prime power factor $N_j = p_j^{v_j}$) whose results can be reassembled without twiddle factors. The separate factors p_j within each N_j must then be dealt with by another algorithm (*e.g.* Cooley–Tukey, which does require twiddle factors). Thus, the DFT on a prime number of points remains undecomposable.

1.3.3.2.3. The Rader algorithm

The previous two algorithms essentially reduce the calculation of the DFT on N points for N composite to the calculation of smaller DFTs on prime numbers of points, the latter remaining irreducible. However, Rader (1968) showed that the p-point DFT for p an odd prime can itself be factored by invoking some extra arithmetic structure present in $\mathbb{Z}/p\mathbb{Z}$.

1.3.3.2.3.1. N an odd prime

The ring $\mathbb{Z}/p\mathbb{Z} = \{0, 1, 2, \ldots, p-1\}$ has the property that its $p-1$ nonzero elements, called *units*, form a *multiplicative group* $U(p)$. In particular, all units $r \in U(p)$ have a unique multiplicative inverse in $\mathbb{Z}/p\mathbb{Z}$, *i.e.* a unit $s \in U(p)$ such that $rs \equiv 1 \bmod p$. This endows $\mathbb{Z}/p\mathbb{Z}$ with the structure of a *finite field*.

Furthermore, $U(p)$ is a *cyclic* group, *i.e.* consists of the successive powers $g^m \bmod p$ of a generator g called a *primitive root mod p* (such a g may not be unique, but it always exists). For instance, for $p = 7$, $U(7) = \{1, 2, 3, 4, 5, 6\}$ is generated by $g = 3$, whose successive powers mod 7 are:

$$g^0 = 1, \quad g^1 = 3, \quad g^2 = 2, \quad g^3 = 6, \quad g^4 = 4, \quad g^5 = 5$$

[see Apostol (1976), Chapter 10].

The basis of Rader's algorithm is to bring to light a hidden regularity in the matrix $F(p)$ by permuting the basis vectors $\mathbf{u}_k$ and $\mathbf{v}_{k^*}$ of $L(\mathbb{Z}/p\mathbb{Z})$ as follows:

$$\begin{aligned}
\mathbf{u}'_0 &= \mathbf{u}_0 \\
\mathbf{u}'_m &= \mathbf{u}_k \quad \text{with } k = g^m, \qquad m = 1, \ldots, p-1; \\
\mathbf{v}'_0 &= \mathbf{v}_0 \\
\mathbf{v}'_{m^*} &= \mathbf{v}_{k^*} \quad \text{with } k^* = g^{m^*}, \quad m^* = 1, \ldots, p-1;
\end{aligned}$$

where g is a primitive root mod p.

With respect to these new bases, the matrix representing $\bar{F}(p)$ will have the following elements:

$$\begin{aligned}
\text{element } (0,0) &= 1 \\
\text{element } (0, m+1) &= 1 \quad \text{for all } m = 0, \ldots p-2, \\
\text{element } (m^*+1, 0) &= 1 \quad \text{for all } m^* = 0, \ldots, p-2, \\
\text{element } (m^*+1, m+1) &= e\left(\frac{k^*k}{p}\right) \\
&= e(g^{(m^*+m)/p}) \\
&\qquad \text{for all } m^* = 0, \ldots, p-2.
\end{aligned}$$

Thus the 'core' $\bar{C}(p)$ of matrix $\bar{F}(p)$, of size $(p-1) \times (p-1)$, formed by the elements with two nonzero indices, has a so-called *skew-circulant* structure because element (m^*, m) depends only on $m^* + m$. Simplification may now occur because multiplication by $\bar{C}(p)$ is closely related to a *cyclic convolution*. Introducing the notation $C(m) = e(g^{m/p})$ we may write the relation $\mathbf{Y}^* = \bar{F}(p)\mathbf{Y}$ in the permuted bases as

$$Y^*(0) = \sum_k Y(k)$$

$$\begin{aligned} Y^*(m^*+1) &= Y(0) + \sum_{m=0}^{p-2} C(m^*+m)Y(m+1) \\ &= Y(0) + \sum_{m=0}^{p-2} C(m^*-m)Z(m) \\ &= Y(0) + (\mathbf{C} * \mathbf{Z})(m^*), \quad m^* = 0, \ldots, p-2, \end{aligned}$$

where $\mathbf{Z}$ is defined by $Z(m) = Y(p-m-2)$, $m = 0, \ldots, p-2$.

Thus $\mathbf{Y}^*$ may be obtained by cyclic convolution of $\mathbf{C}$ and $\mathbf{Z}$, which may for instance be calculated by

$$\mathbf{C} * \mathbf{Z} = F(p-1)[\bar{F}(p-1)[\mathbf{C}] \times \bar{F}(p-1)[\mathbf{Z}]],$$

where $\times$ denotes the component-wise multiplication of vectors. Since p is odd, $p-1$ is always divisible by 2 and may even be highly composite. In that case, factoring $\bar{F}(p-1)$ by means of the Cooley–Tukey or Good methods leads to an algorithm of complexity $p \log p$ rather than p^2 for $\bar{F}(p)$. An added bonus is that, because $g^{(p-1)/2} = -1$, the elements of $\bar{F}(p-1)[\mathbf{C}]$ can be shown to be either purely real or purely imaginary, which halves the number of real multiplications involved.

1.3.3.2.3.2. N a power of an odd prime

This idea was extended by Winograd (1976, 1978) to the treatment of prime powers $N = p^\nu$, using the cyclic structure of the multiplicative group of units $U(p^\nu)$. The latter consists of all those elements of $\mathbb{Z}/p^\nu\mathbb{Z}$ which are not divisible by p, and thus has $q_\nu = p^{\nu-1}(p-1)$ elements. It is cyclic, and there exist primitive roots g modulo p^ν such that

$$U(p^\nu) = \{1, g, g^2, g^3, \ldots, g^{q_\nu - 1}\}.$$

The $p^{\nu-1}$ elements divisible by p, which are divisors of zero, have to be treated separately just as 0 had to be treated separately for $N = p$.

When $k^* \notin U(p^\nu)$, then $k^* = pk_1^*$ with $k_1^* \in \mathbb{Z}/p^{\nu-1}\mathbb{Z}$. The results $X^*(pk_1^*)$ are p-decimated, hence can be obtained *via* the $p^{\nu-1}$-point DFT of the $p^{\nu-1}$-periodized data $\mathbf{Y}$:

$$X^*(pk_1^*) = \bar{F}(p^{\nu-1})[\mathbf{Y}](k_1^*)$$

with

$$Y(k_1) = \sum_{k_2 \in \mathbb{Z}/p\mathbb{Z}} X(k_1 + p^{\nu-1}k_2).$$

When $k^* \in U(p^\nu)$, then we may write

$$X^*(k^*) = X_0^*(k^*) + X_1^*(k^*),$$

where $\mathbf{X}_0^*$ contains the contributions from $k \notin U(p^\nu)$ and $\mathbf{X}_1^*$ those from $k \in U(p^\nu)$. By a converse of the previous calculation, $\mathbf{X}_0^*$ arises from p-decimated data $\mathbf{Z}$, hence is the $p^{\nu-1}$-periodization of the $p^{\nu-1}$-point DFT of these data:

$$X_0^*(p^{\nu-1}k_1^* + k_2^*) = \bar{F}(p^{\nu-1})[\mathbf{Z}](k_2^*)$$

with

$$Z(k_2) = X(pk_2), \qquad k_2 \in \mathbb{Z}/p^{\nu-1}\mathbb{Z}$$

(the $p^{\nu-1}$-periodicity follows implicit from the fact that the transform on the right-hand side is independent of $k_1^* \in \mathbb{Z}/p\mathbb{Z}$).

Finally, the contribution X_1^* from all $k \in U(p^\nu)$ may be calculated by reindexing by the powers of a primitive root g modulo p^ν, *i.e.* by writing

$$X_1^*(g^{m^*}) = \sum_{m=0}^{q_\nu - 1} X(g^m) e(g^{(m^*+m)/p^\nu})$$

then carrying out the multiplication by the skew-circulant matrix core as a convolution.

Thus the DFT of size p^ν may be reduced to two DFTs of size $p^{\nu-1}$ (dealing, respectively, with p-decimated results and p-decimated data) and a convolution of size $q_\nu = p^{\nu-1}(p-1)$. The latter may be 'diagonalized' into a multiplication by purely real or purely imaginary numbers (because $g^{(q_\nu/2)} = -1$) by two DFTs, whose factoring in turn leads to DFTs of size $p^{\nu-1}$ and $p-1$. This method, applied recursively, allows the complete decomposition of the DFT on p^ν points into arbitrarily small DFTs.

1.3.3.2.3.3. N a power of 2

When $N = 2^\nu$, the same method can be applied, except for a slight modification in the calculation of $\mathbf{X}_1^*$. There is no primitive root modulo 2^ν for $\nu > 2$: the group $U(2^\nu)$ is the direct product of *two* cyclic groups, the first (of order 2) generated by -1, the second (of order $N/4$) generated by 3 or 5. One then uses a representation

$$\begin{aligned} k &= (-1)^{m_1} 5^{m_2} \\ k^* &= (-1)^{m_1^*} 5^{m_2^*} \end{aligned}$$

and the reindexed core matrix gives rise to a two-dimensional convolution. The latter may be carried out by means of two 2D DFTs on $2 \times (N/4)$ points.

1.3.3.2.4. The Winograd algorithms

The cyclic convolutions generated by Rader's multiplicative reindexing may be evaluated more economically than through DFTs if they are re-examined within a new algebraic setting, namely the theory of congruence classes of polynomials [see, for instance, Blahut (1985), Chapter 2; Schroeder (1986), Chapter 24].

The set, denoted $\mathbb{K}[X]$, of polynomials in one variable with coefficients in a given field $\mathbb{K}$ has many of the formal properties of the set $\mathbb{Z}$ of rational integers: it is a *ring* with no zero divisors and has a *Euclidean algorithm* on which a theory of divisibility can be built.

Given a polynomial $P(z)$, then for every $W(z)$ there exist unique polynomials $Q(z)$ and $R(z)$ such that

$$W(z) = P(z)Q(z) + R(z)$$

and

$$\text{degree } (R) < \text{degree } (P).$$

$R(z)$ is called the *residue* of $H(z)$ modulo $P(z)$. Two polynomials $H_1(z)$ and $H_2(z)$ having the same residue modulo $P(z)$ are said to be *congruent* modulo $P(z)$, which is denoted by

$$H_1(z) \equiv H_2(z) \bmod P(z).$$

If $H(z) \equiv 0 \bmod P(z)$, $H(z)$ is said to be *divisible* by $P(z)$. If $H(z)$ only has divisors of degree zero in $\mathbb{K}[X]$, it is said to be *irreducible over* $\mathbb{K}$ (this notion depends on $\mathbb{K}$). Irreducible polynomials play in $\mathbb{K}[X]$ a role analogous to that of prime numbers in $\mathbb{Z}$, and any polynomial over $\mathbb{K}$ has an essentially unique factorization as a product of irreducible polynomials.

There exists a *Chinese remainder theorem* (CRT) for polynomials. Let $P(z) = P_1(z) \ldots P_d(z)$ be factored into a product of pairwise coprime polynomials [*i.e.* $P_i(z)$ and $P_j(z)$ have no common factor for $i \neq j$]. Then the system of congruence equations

$$H(z) \equiv H_j(z) \bmod P_j(z), \quad j = 1, \ldots, d,$$

has a unique solution $H(z)$ modulo $P(z)$. This solution may be constructed by a procedure similar to that used for integers. Let

$$Q_j(z) = P(z)/P_j(z) = \prod_{i \neq j} P_i(z).$$

Then P_j and Q_j are coprime, and the Euclidean algorithm may be used to obtain polynomials $p_j(z)$ and $q_j(z)$ such that

$$p_j(z)P_j(z) + q_j(z)Q_j(z) = 1.$$

With $S_i(z) = q_i(z)Q_i(z)$, the polynomial

$$H(z) = \sum_{i=1}^{d} S_i(z)H_i(z) \bmod P(z)$$

is easily shown to be the desired solution.

As with integers, it can be shown that the 1:1 correspondence between $H(z)$ and $H_j(z)$ sends sums to sums and products to products, *i.e.* establishes a *ring isomorphism*:

$$\mathbb{K}[X] \bmod P \cong (\mathbb{K}[X] \bmod P_1) \times \ldots \times (\mathbb{K}[X] \bmod P_d).$$

These results will now be applied to the efficient calculation of cyclic convolutions. Let $\mathbf{U} = (u_0, u_1, \ldots, u_{N-1})$ and $\mathbf{V} = (v_0, v_1, \ldots, v_{N-1})$ be two vectors of length N, and let $\mathbf{W} = (w_0, w_1, \ldots, w_{N-1})$ be obtained by cyclic convolution of $\mathbf{U}$ and $\mathbf{V}$:

$$w_n = \sum_{m=0}^{N-1} u_m v_{n-m}, \quad n = 0, \ldots, N-1.$$

The very simple but crucial result is that this cyclic convolution may be carried out by *polynomial multiplication modulo* $(z^N - 1)$: if

$$U(z) = \sum_{l=0}^{N-1} u_l z^l$$
$$V(z) = \sum_{m=0}^{N-1} v_m z^m$$
$$W(z) = \sum_{n=0}^{N-1} w_n z^n$$

then the above relation is equivalent to

$$W(z) \equiv U(z)V(z) \bmod (z^N - 1).$$

Now the polynomial $z^N - 1$ can be *factored* over the field of rational numbers into irreducible factors called *cyclotomic polynomials*: if d is the number of divisors of N, including 1 and N, then

$$z^N - 1 = \prod_{i=1}^{d} P_i(z),$$

where the cyclotomics $P_i(z)$ are well known (Nussbaumer, 1981; Schroeder, 1986, Chapter 22). We may now invoke the CRT, and exploit the ring isomorphism it establishes to simplify the calculation of $W(z)$ from $U(z)$ and $V(z)$ as follows:

(i) compute the d residual polynomials

$$U_i(z) \equiv U(z) \bmod P_i(z), \quad i = 1, \ldots, d,$$
$$V_i(z) \equiv V(z) \bmod P_i(z), \quad i = 1, \ldots, d;$$

(ii) compute the d polynomial products

$$W_i(z) \equiv U_i(z)V_i(z) \bmod P_i(z), \quad i = 1, \ldots, d;$$

(iii) use the CRT reconstruction formula just proved to recover $W(z)$ from the $W_i(z)$:

$$W(z) \equiv \sum_{i=1}^{d} S_i(z)W_i(z) \bmod (z^N - 1).$$

When N is not too large, *i.e.* for 'short cyclic convolutions', the $P_i(z)$ are very simple, with coefficients 0 or ± 1, so that (i) only involves a small number of additions. Furthermore, special techniques have been developed to multiply general polynomials modulo cyclotomic polynomials, thus helping keep the number of multiplications in (ii) and (iii) to a minimum. As a result, cyclic convolutions can be calculated rapidly when N is sufficiently composite.

It will be recalled that Rader's multiplicative indexing often gives rise to cyclic convolutions of length $p - 1$ for p an odd prime. Since $p - 1$ is highly composite for all $p \leq 50$ other than 23 and 47, these cyclic convolutions can be performed more efficiently by the above procedure than by DFT.

These combined algorithms are due to Winograd (1977, 1978, 1980), and are known collectively as 'Winograd small FFT algorithms'. Winograd also showed that they can be thought of as bringing the DFT matrix $\mathbf{F}$ to the following 'normal form':

$$\mathbf{F} = \mathbf{CBA},$$

where

$\mathbf{A}$ is an integer matrix with entries 0, ± 1, defining the 'pre-additions',

$\mathbf{B}$ is a diagonal matrix of multiplications,

$\mathbf{C}$ is a matrix with entries 0, ± 1, $\pm i$, defining the 'post-additions'.

The elements on the diagonal of $\mathbf{B}$ can be shown to be either real or pure imaginary, by the same argument as in Section 1.3.3.2.3.1. Matrices $\mathbf{A}$ and $\mathbf{C}$ may be rectangular rather than square, so that intermediate results may require extra storage space.

1.3.3.3. Multidimensional algorithms

From an algorithmic point of view, the distinction between one-dimensional (1D) and multidimensional DFTs is somewhat blurred by the fact that some factoring techniques turn a 1D transform into a multidimensional one. The distinction made here, however, is a practical one and is based on the dimensionality of the indexing sets for data and results. This section will therefore be concerned with the problem of factoring the DFT when the *indexing sets* for the input data and output results are multidimensional.

1.3.3.3.1. The method of successive one-dimensional transforms

The DFT was defined in Section 1.3.2.7.4 in an n-dimensional setting and it was shown that when the decimation matrix $\mathbf{N}$ is diagonal, say $\mathbf{N} = \mathrm{diag}(N^{(1)}, N^{(2)}, \ldots, N^{(n)})$, then $\bar{F}(N)$ has a tensor product structure:

$$\bar{F}(\mathbf{N}) = \bar{F}(N^{(1)}) \otimes \bar{F}(N^{(2)}) \otimes \ldots \otimes \bar{F}(N^{(n)}).$$

This may be rewritten as follows:

$$\begin{aligned}\bar{F}(\mathbf{N}) = {} & [\bar{F}(N^{(1)}) \otimes I_{N^{(2)}} \otimes \ldots \otimes I_{N^{(n)}}] \\ & \times [I_{N^{(1)}} \otimes \bar{F}(N^{(2)}) \otimes \ldots \otimes I_{N^{(n)}}] \\ & \times \ldots \\ & \times [I_{N^{(1)}} \otimes I_{N^{(2)}} \otimes \ldots \otimes \bar{F}(N^{(n)}],\end{aligned}$$

where the I's are identity matrices and $\times$ denotes ordinary matrix multiplication. The matrix within each bracket represents a one-dimensional DFT along one of the n dimensions, the other dimensions being left untransformed. As these matrices commute, the order in which the successive 1D DFTs are performed is immaterial.

This is the most straightforward method for building an n-dimensional algorithm from existing 1D algorithms. It is known in crystallography under the name of 'Beevers–Lipson factorization' (Section 1.3.4.3.1), and in signal processing as the 'row–column method'.

1.3.3.3.2. Multidimensional factorization

Substantial reductions in the arithmetic cost, as well as gains in flexibility, can be obtained if the factoring of the DFT is carried out in several dimensions simultaneously. The presentation given here is a generalization of that of Mersereau & Speake (1981), using the abstract setting established independently by Auslander, Tolimieri & Winograd (1982).

Let us return to the general n-dimensional setting of Section 1.3.2.7.4, where the DFT was defined for an arbitrary decimation matrix $\mathbf{N}$ by the formulae (where $|\mathbf{N}|$ denotes $|\det \mathbf{N}|$):

$$\begin{aligned} F(\mathbf{N}): \quad X(\mathbf{k}) &= \frac{1}{|\mathbf{N}|} \sum_{\mathbf{k}^*} X^*(\mathbf{k}^*) e[-\mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})] \\ \bar{F}(\mathbf{N}): \quad X^*(\mathbf{k}^*) &= \sum_{\mathbf{k}} X(\mathbf{k}) e[\mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})] \end{aligned}$$

with

$$\mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n, \quad \mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n.$$

1.3.3.3.2.1. Multidimensional Cooley–Tukey factorization

Let us now assume that this decimation can be factored into d successive decimations, *i.e.* that

$$\mathbf{N} = \mathbf{N}_1\mathbf{N}_2 \ldots \mathbf{N}_{d-1}\mathbf{N}_d$$

and hence

$$\mathbf{N}^T = \mathbf{N}_d^T\mathbf{N}_{d-1}^T \ldots \mathbf{N}_2^T\mathbf{N}_1^T.$$

Then the coset decomposition formulae corresponding to these successive decimations (Section 1.3.2.7.1) can be combined as follows:

$$\begin{aligned}\mathbb{Z}^n &= \bigcup_{\mathbf{k}_1} (\mathbf{k}_1 + \mathbf{N}_1\mathbb{Z}^n) \\ &= \bigcup_{\mathbf{k}_1} \left\{ \mathbf{k}_1 + \mathbf{N}_1 \left[\bigcup_{\mathbf{k}_2} (\mathbf{k}_2 + \mathbf{N}_2\mathbb{Z}^n) \right] \right\} \\ &= \ldots \\ &= \bigcup_{\mathbf{k}_1} \ldots \bigcup_{\mathbf{k}_d} (\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d + \mathbf{N}\mathbb{Z}^n)\end{aligned}$$

with $\mathbf{k}_j \in \mathbb{Z}^n/\mathbf{N}_j\mathbb{Z}^n$. Therefore, any $\mathbf{k} \in \mathbb{Z}/\mathbf{N}\mathbb{Z}^n$ may be written uniquely as

$$\mathbf{k} = \mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d.$$

Similarly:

$$\begin{aligned}\mathbb{Z}^n &= \bigcup_{\mathbf{k}_d^*} (\mathbf{k}_d^* + \mathbf{N}_d^T\mathbb{Z}^n) \\ &= \ldots \\ &= \bigcup_{\mathbf{k}_d^*} \ldots \bigcup_{\mathbf{k}_1^*} (\mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^* \\ &\quad + \mathbf{N}^T\mathbb{Z}^n)\end{aligned}$$

so that any $\mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ may be written uniquely as

$$\mathbf{k}^* = \mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^*$$

with $\mathbf{k}_j^* \in \mathbb{Z}^n/\mathbf{N}_j^T\mathbb{Z}^n$. These decompositions are the vector analogues of the multi-radix number representation systems used in the Cooley–Tukey factorization.

We may then write the definition of $\bar{F}(\mathbf{N})$ with $d = 2$ factors as

$$\begin{aligned}X^*(\mathbf{k}_2^* + \mathbf{N}_2^T\mathbf{k}_1^*) = {} & \sum_{\mathbf{k}_1} \sum_{\mathbf{k}_2} X(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2) \\ & \times e[(\mathbf{k}_2^{*T} + \mathbf{k}_1^{*T}\mathbf{N}_2)\mathbf{N}_2^{-1}\mathbf{N}_1^{-1}(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2)].\end{aligned}$$

The argument of $e(-)$ may be expanded as

$$\mathbf{k}_2^* \cdot (\mathbf{N}^{-1}\mathbf{k}_1) + \mathbf{k}_1^* \cdot (\mathbf{N}_1^{-1}\mathbf{k}_1) + \mathbf{k}_2^* \cdot (\mathbf{N}_2^{-1}\mathbf{k}_2) + \mathbf{k}_1^* \cdot \mathbf{k}_2.$$

The first summand may be recognized as a twiddle factor, the second and third as the kernels of $\bar{F}(\mathbf{N}_1)$ and $\bar{F}(\mathbf{N}_2)$, respectively, while the fourth is an integer which may be dropped. We are thus led to a 'vector-radix' version of the Cooley–Tukey algorithm, in which the successive decimations may be introduced in all n dimensions simultaneously by general integer matrices. The computation may be decomposed into five stages analogous to those of the one-dimensional algorithm of Section 1.3.3.2.1:

(i) form the $|\mathbf{N}_1|$ vectors $\mathbf{Y}_{\mathbf{k}_1}$ of shape $\mathbf{N}_2$ by

$$Y_{\mathbf{k}_1}(\mathbf{k}_2) = X(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n, \quad \mathbf{k}_2 \in \mathbb{Z}^n/\mathbf{N}_2\mathbb{Z}^n;$$

(ii) calculate the $|\mathbf{N}_1|$ transforms $\mathbf{Y}^*_{\mathbf{k}_1}$ on $|\mathbf{N}_2|$ points:

$$Y^*_{\mathbf{k}_1}(\mathbf{k}_2^*) = \sum_{\mathbf{k}_2} e[\mathbf{k}_2^* \cdot (\mathbf{N}_2^{-1}\mathbf{k}_2)] Y_{\mathbf{k}_1}(\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n;$$

(iii) form the $|\mathbf{N}_2|$ vectors $\mathbf{Z}_{\mathbf{k}_2^*}$ of shape $\mathbf{N}_1$ by

$$Z_{\mathbf{k}_2^*}(\mathbf{k}_1) = e[\mathbf{k}_2^* \cdot (\mathbf{N}^{-1}\mathbf{k}_1)] Y^*_{\mathbf{k}_1}(\mathbf{k}_2^*), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n, \quad \mathbf{k}_2^* \in \mathbb{Z}^n/\mathbf{N}_2^T\mathbb{Z}^n;$$

(iv) calculate the $|\mathbf{N}_2|$ transforms $\mathbf{Z}^*_{\mathbf{k}_2^*}$ on $|\mathbf{N}_1|$ points:

$$Z^*_{\mathbf{k}_2^*}(\mathbf{k}_1^*) = \sum_{\mathbf{k}_1} e[\mathbf{k}_1^* \cdot (\mathbf{N}_1^{-1}\mathbf{k}_1)] Z_{\mathbf{k}_2^*}(\mathbf{k}_1), \quad \mathbf{k}_2^* \in \mathbb{Z}^n/\mathbf{N}_2^T\mathbb{Z}^n;$$

(v) collect $X^*(\mathbf{k}_2^* + \mathbf{N}_2^T\mathbf{k}_1^*)$ as $Z^*_{\mathbf{k}_2^*}(\mathbf{k}_1^*)$.

The initial $|\mathbf{N}|$-point transform $\bar{F}(\mathbf{N})$ can thus be performed as $|\mathbf{N}_1|$ transforms $\bar{F}(\mathbf{N}_2)$ on $|\mathbf{N}_2|$ points, followed by $|\mathbf{N}_2|$ transforms $\bar{F}(\mathbf{N}_1)$ on $|\mathbf{N}_1|$ points. This process can be applied successively to all d factors. The same decomposition applies to $F(\mathbf{N})$, up to the complex conjugation of twiddle factors, the normalization factor $1/|\mathbf{N}|$ being obtained as the product of the factors $1/|\mathbf{N}_j|$ in the successive partial transforms $F(\mathbf{N}_j)$.

The geometric interpretation of this factorization in terms of partial transforms on translates of sublattices applies in full to this n-dimensional setting; in particular, the twiddle factors are seen to be related to the residual translations which place the sublattices in register within the big lattice. If the intermediate transforms are performed *in place*, then the quantity

$$X^*(\mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T\mathbf{N}_{d-1}^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^*)$$

will eventually be found at location

$$\mathbf{k}_1^* + \mathbf{N}_1\mathbf{k}_2^* + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d^*,$$

so that the final results will have to be *unscrambled* by a process which may be called 'coset reversal', the vector equivalent of digit reversal.

Factoring by 2 in all n dimensions simultaneously, *i.e.* taking $\mathbf{N} = 2\mathbf{M}$, leads to 'n-dimensional butterflies'. Decimation in time corresponds to the choice $\mathbf{N}_1 = 2\mathbf{I}$, $\mathbf{N}_2 = \mathbf{M}$, so that $\mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ is an n-dimensional parity class; the calculation then proceeds by

$$\begin{aligned} Y_{\mathbf{k}_1}(\mathbf{k}_2) = X(\mathbf{k}_1 + 2\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n, \quad \mathbf{k}_2 \in \mathbb{Z}^n/\mathbf{M}\mathbb{Z}^n, \\ Y^*_{\mathbf{k}_1} = \bar{F}(\mathbf{M})[\mathbf{Y}_{\mathbf{k}_1}], \quad \mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n; \\ X^*(\mathbf{k}_2^* + \mathbf{M}^T\mathbf{k}_1^*) = \sum_{\mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n} (-1)^{\mathbf{k}_1^* \cdot \mathbf{k}_1} \\ \times e[\mathbf{k}_2^* \cdot (\mathbf{N}^{-1}\mathbf{k}_1)] Y^*_{\mathbf{k}_1}(\mathbf{k}_2^*). \end{aligned}$$

Decimation in frequency corresponds to the choice $\mathbf{N}_1 = \mathbf{M}$, $\mathbf{N}_2 = 2\mathbf{I}$, so that $\mathbf{k}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n$ labels 'octant' blocks of shape $\mathbf{M}$; the calculation then proceeds through the following steps:

$$\begin{aligned} Z_{\mathbf{k}_2^*}(\mathbf{k}_1) = \left[\sum_{\mathbf{k}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n} (-1)^{\mathbf{k}_2^* \cdot \mathbf{k}_2} X(\mathbf{k}_1 + \mathbf{M}\mathbf{k}_2) \right] \\ \times e[\mathbf{k}_2^* \cdot (\mathbf{N}^{-1}\mathbf{k}_1)], \\ \mathbf{Z}^*_{\mathbf{k}_2^*} = \bar{F}(\mathbf{M})[\mathbf{Z}_{\mathbf{k}_2^*}], \\ X^*(\mathbf{k}_2^* + 2\mathbf{k}_1^*) = Z^*_{\mathbf{k}_2^*}(\mathbf{k}_1^*), \end{aligned}$$

i.e. the 2^n parity classes of results, corresponding to the different $\mathbf{k}_2^* \in \mathbb{Z}^n/2\mathbb{Z}^n$, are obtained separately. When the dimension n is 2 and the decimating matrix is diagonal, this analysis reduces to the 'vector radix FFT' algorithms proposed by Rivard (1977) and Harris *et al.* (1977). These lead to substantial reductions in the number M of multiplications compared to the row–column method: M is reduced to $3M/4$ by simultaneous 2×2 factoring, and to $15M/32$ by simultaneous 4×4 factoring.

The use of a nondiagonal decimating matrix may bring savings in computing time if the spectrum of the band-limited function under study is of such a shape as to pack more compactly in a nonrectangular than in a rectangular lattice (Mersereau, 1979). If, for instance, the support K of the spectrum Φ is contained in a sphere, then a decimation matrix producing a close packing of these spheres will yield an aliasing-free DFT algorithm with fewer sample points than the standard algorithm using a rectangular lattice.

1.3.3.3.2.2. Multidimensional prime factor algorithm

Suppose that the decimation matrix $\mathbf{N}$ is diagonal

$$\mathbf{N} = \text{diag } (N^{(1)}, N^{(2)}, \ldots, N^{(n)})$$

and let each diagonal element be written in terms of its prime factors:

$$N^{(i)} = \prod_{j=1}^{m} p_j^{\nu(i,\,j)},$$

where m is the total number of distinct prime factors present in the $N^{(i)}$.

The CRT may be used to turn each 1D transform along dimension i ($i = 1, \ldots, n$) into a multidimensional transform with a separate 'pseudo-dimension' for each distinct prime factor of $N^{(i)}$; the number μ_i, of these pseudo-dimensions is equal to the cardinality of the set:

$$\{j \in \{1, \ldots, m\} | \nu(i, j) > 0 \text{ for some } i\}.$$

The full n-dimensional transform thus becomes μ-dimensional, with $\mu = \sum_{i=1}^{n} \mu_i$.

We may now permute the μ pseudo-dimensions so as to bring into contiguous position those corresponding to the same prime factor p_j; the m resulting groups of pseudo-dimensions are said to define 'p-primary' blocks. The initial transform is now written as a tensor product of m p-primary transforms, where transform j is on

$$p_j^{\nu(1,\,j)} \times p_j^{\nu(2,\,j)} \times \ldots \times p_j^{\nu(n,\,j)}$$

points [by convention, dimension i is not transformed if $\nu(i, j) = 0$]. These p-primary transforms may be computed, for instance, by multidimensional Cooley–Tukey factorization (Section 1.3.3.3.1), which is faster than the straightforward row–column method. The final results may then be obtained by reversing all the permutations used.

The extra gain with respect to the multidimensional Cooley–Tukey method is that *there are no twiddle factors between p-primary pieces corresponding to different primes p*.

The case where **N** is not diagonal has been examined by Guessoum & Mersereau (1986).

1.3.3.3.2.3. Nesting of Winograd small FFTs

Suppose that the CRT has been used as above to map an n-dimensional DFT to a μ-dimensional DFT. For each $\kappa = 1, \ldots, \mu$ [κ runs over those pairs (i, j) such that $\nu(i, j) > 0$], the Rader/Winograd procedure may be applied to put the matrix of the κth 1D DFT in the **CBA** normal form of a Winograd small FFT. The full DFT matrix may then be written, up to permutation of data and results, as

$$\bigotimes_{\kappa=1}^{\mu} (\mathbf{C}_\kappa \mathbf{B}_\kappa \mathbf{A}_\kappa).$$

A well known property of the tensor product of matrices allows this to be rewritten as

$$\left(\bigotimes_{\gamma=1}^{\mu} \mathbf{C}_\gamma\right) \times \left(\bigotimes_{\beta=1}^{\mu} \mathbf{B}_\beta\right) \times \left(\bigotimes_{\alpha=1}^{\mu} \mathbf{A}_\alpha\right)$$

and thus to form a matrix in which the *combined* pre-addition, multiplication and post-addition matrices have been *precomputed*. This procedure, called *nesting*, can be shown to afford a reduction of the arithmetic operation count compared to the row–column method (Morris, 1978).

Clearly, the nesting rearrangement need not be applied to all μ dimensions, but can be restricted to any desired subset of them.

1.3.3.3.2.4. The Nussbaumer–Quandalle algorithm

Nussbaumer's approach views the DFT as the evaluation of certain polynomials constructed from the data (as in Section 1.3.3.2.4). For instance, putting $\omega = e(1/N)$, the 1D N-point DFT

$$X^*(k^*) = \sum_{k=0}^{N-1} X(k)\omega^{k^* k}$$

may be written

$$X^*(k^*) = Q(\omega^{k^*}),$$

where the polynomial Q is defined by

$$Q(z) = \sum_{k=0}^{N-1} X(k) z^k.$$

Let us consider (Nussbaumer & Quandalle, 1979) a 2D transform of size $N \times N$:

$$X^*(k_1^*, k_2^*) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} X(k_1, k_2)\omega^{k_1^* k_1 + k_2^* k_2}.$$

By introduction of the polynomials

$$Q_{k_2}(z) = \sum_{k_1} X(k_1, k_2) z^{k_1}$$

$$R_{k_2^*}(z) = \sum_{k_2} \omega^{k_2^* k_2} Q_{k_2}(z),$$

this may be rewritten:

$$X^*(k_1^*, k_2^*) = R_{k_2^*}(\omega^{k_1^*}) = \sum_{k_2} \omega^{k_2^* k_2} Q_{k_2}(\omega^{k_1^*}).$$

Let us now suppose that k_1^* is coprime to N. Then k_1^* has a unique inverse modulo N (denoted by $1/k_1^*$), so that multiplication by k_1^* simply permutes the elements of $\mathbb{Z}/N\mathbb{Z}$ and hence

$$\sum_{k_2=0}^{N-1} f(k_2) = \sum_{k_2=0}^{N-1} f(k_1^* k_2)$$

for any function f over $\mathbb{Z}/N\mathbb{Z}$. We may thus write:

$$\begin{aligned} X^*(k_1^*, k_2^*) &= \sum_{k_2} \omega^{k_1^* k_2^* k_2} Q_{k_1^* k_2}(\omega^{k_1^*}) \\ &= S_{k_1^* k_2}(\omega^{k_1^*}) \end{aligned}$$

where

$$S_{k^*}(z) = \sum_{k_2} z^{k^* k_2} Q_{k_2}(z).$$

Since only the value of polynomial $S_{k^*}(z)$ at $z = \omega^{k_1^*}$ is involved in the result, the computation of S_{k^*} may be carried out modulo the unique cyclotomic polynomial $P(z)$ such that $P(\omega^{k_1^*}) = 0$. Thus, if we define:

$$T_{k^*}(z) = \sum_{k_2} z^{k^* k_2} Q_{k_2}(z) \bmod P(z)$$

we may write:

$$X^*(k_1^*, k_2^*) = T_{k_1^* k_2^*}(\omega^{k_1^*})$$

or equivalently

$$X^*\left(k_1^*, \frac{k_2^*}{k_1^*}\right) = T_{k_2^*}(\omega^{k_1^*}).$$

For N an odd prime p, all nonzero values of k_1^* are coprime with p so that the $p \times p$-point DFT may be calculated as follows:

(1) form the polynomials

$$T_{k_2^*}(z) = \sum_{k_1} \sum_{k_2} X(k_1, k_2) z^{k_1 + k_2^* k_2} \bmod P(z)$$

for $k_2^* = 0, \ldots, p-1$;

(2) evaluate $T_{k_2^*}(\omega^{k_1^*})$ for $k_1^* = 0, \ldots, p-1$;

(3) put $X^*(k_1^*, k_2^*/k_1^*) = T_{k_2^*}(\omega^{k_1^*})$;

(4) calculate the terms for $k_1^* = 0$ separately by

$$X^*(0, k_2^*) = \sum_{k_2} \left[\sum_{k_1} X(k_1, k_2) \right] \omega^{k_2^* k_2}.$$

Step (1) is a set of p 'polynomial transforms' involving no multiplications; step (2) consists of p DFTs on p points each since if

$$T_{k_2^*}(z) = \sum_{k_1} Y_{k_2^*}(k_1) z^{k_1}$$

then

$$T_{k_2^*}(\omega^{k_1^*}) = \sum_{k_1} Y_{k_2^*}(k_1) \omega^{k_1^* k_1} = Y_{k_2^*}^*(k_1^*);$$

step (3) is a permutation; and step (4) is a p-point DFT. Thus the 2D DFT on $p \times p$ points, which takes $2p$ p-point DFTs by the row–column method, involves only $(p+1)$ p-point DFTs; the other DFTs have been replaced by polynomial transforms involving only additions.

This procedure can be extended to n dimensions, and reduces the number of 1D p-point DFTs from np^{n-1} for the row–column method to $(p^n - 1)/(p-1)$, at the cost of introducing extra additions in the polynomial transforms.

A similar algorithm has been formulated by Auslander *et al.* (1983) in terms of Galois theory.

1.3.3.3.3. Global algorithm design

1.3.3.3.3.1. From local pieces to global algorithms

The mathematical analysis of the structure of DFT computations has brought to light a broad variety of possibilities for reducing or reshaping their arithmetic complexity. All of them are 'analytic' in that they break down large transforms into a succession of smaller ones.

These results may now be considered from the converse 'synthetic' viewpoint as providing a list of procedures for assembling them:

(i) the building blocks are one-dimensional p-point algorithms for p a small prime;

(ii) the low-level connectors are the multiplicative reindexing methods of Rader and Winograd, or the polynomial transform reindexing method of Nussbaumer and Quandalle, which allow the construction of efficient algorithms for larger primes p, for prime powers p^ν, and for p-primary pieces of shape $p^\nu \times \ldots \times p^\nu$;

(iii) the high-level connectors are the additive reindexing scheme of Cooley–Tukey, the Chinese remainder theorem reindexing, and the tensor product construction;

(iv) nesting may be viewed as the 'glue' which seals all elements.

The simplest DFT may then be carried out into a global algorithm in many different ways. The diagrams in Fig. 1.3.3.1 illustrate a few of the options available to compute a 400-point DFT. They may differ greatly in their arithmetic operation counts.

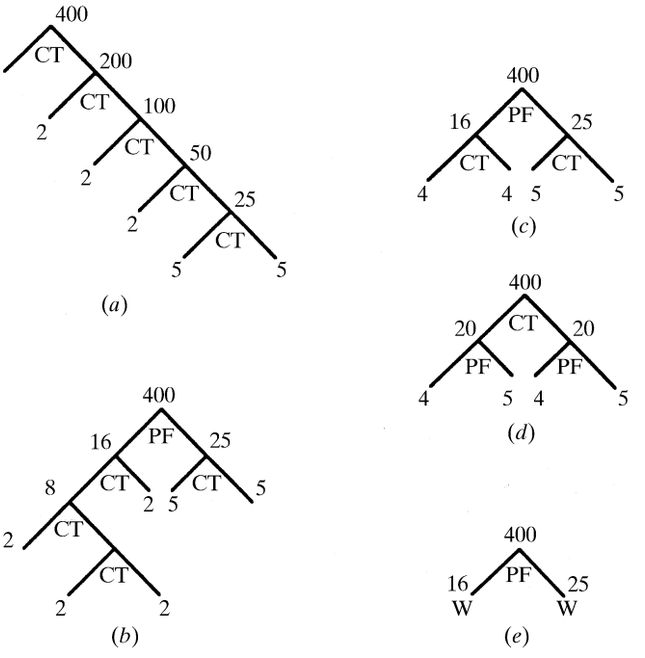


Fig. 1.3.3.1. A few global algorithms for computing a 400-point DFT. CT: Cooley–Tukey factorization. PF: prime factor (or Good) factorization. W: Winograd algorithm.

1.3.3.3.3.2. Computer architecture considerations

To obtain a truly useful measure of the computational complexity of a DFT algorithm, its arithmetic operation count must be tempered by computer architecture considerations. Three main types of trade-offs must be borne in mind:

(i) reductions in floating-point (f.p.) arithmetic count are obtained by reindexing, hence at the cost of an increase in integer arithmetic on addresses, although some shortcuts may be found (Uhrich, 1969; Burrus & Eschenbacher, 1981);

(ii) reduction in the f.p. multiplication count usually leads to a large increase in the f.p. addition count (Morris, 1978);

(iii) nesting can increase execution speed, but causes a loss of modularity and hence complicates program development (Silverman, 1977; Kolba & Parks, 1977).

Many of the mathematical developments above took place in the context of single-processor serial computers, where f.p. addition is substantially cheaper than f.p. multiplication but where integer address arithmetic has to compete with f.p. arithmetic for processor cycles. As a result, the alternatives to the Cooley–Tukey algorithm hardly ever led to particularly favourable trade-offs, thus creating the impression that there was little to gain by switching to more exotic algorithms.

The advent of new machine architectures with vector and/or parallel processing features has greatly altered this picture (Pease, 1968; Korn & Lambiotte, 1979; Fornberg, 1981; Swartzrauber, 1984):

(i) *pipelining* equalizes the cost of f.p. addition and f.p. multiplication, and the ideal 'blend' of the two types of operations depends solely on the number of adder and multiplier units available in each machine;

(ii) integer address arithmetic is delegated to specialized arithmetic and logical units (ALUs) operating concurrently with the f.p. units, so that complex reindexing schemes may be used without loss of overall efficiency.

Another major consideration is that of data flow [see *e.g.* Nawab & McClellan (1979)]. Serial machines only have few registers and few paths connecting them, and allow little or no overlap between computation and data movement. New architectures, on the other hand, comprise banks of vector registers (or 'cache memory') besides the usual internal registers, and dedicated ALUs can service data transfers between several of them simultaneously and concurrently with computation.

In this new context, the devices described in Sections 1.3.3.2 and 1.3.3.3 for altering the balance between the various types of arithmetic operations, and reshaping the data flow during the computation, are invaluable. The field of machine-dependent DFT algorithm design is thriving on them [see *e.g.* Temperton (1983*a*,*b*,*c*, 1985); Agarwal & Cooley (1986, 1987)].

1.3.3.3.3.3. The Johnson–Burrus family of algorithms

In order to explore systematically all possible algorithms for carrying out a given DFT computation, and to pick the one best suited to a given machine, attempts have been made to develop:

(i) a high-level notation of describing all the ingredients of a DFT computation, including data permutation and data flow;

(ii) a formal calculus capable of operating on these descriptions so as to represent all possible reorganizations of the computation;

(iii) an automatic procedure for evaluating the performance of a given algorithm on a specific architecture.

Task (i) can be accomplished by systematic use of a tensor product notation to represent the various stages into which the DFT can be factored (reindexing, small transforms on subsets of indices, twiddle factors, digit-reversal permutations).

Task (ii) may for instance use the Winograd **CBA** normal form for each small transform, then apply the rules governing the rearrangement of tensor product $\bigotimes$ and ordinary product $\times$ operations on matrices. The matching of these rearrangements to the architecture of a vector and/or parallel computer can be formalized algebraically [see *e.g.* Chapter 2 of Tolimieri *et al.* (1989)].

Task (iii) is a complex search which requires techniques such as dynamic programming (Bellman, 1958).

Johnson & Burrus (1983) have proposed and tested such a scheme to identify the optimal trade-offs between prime factor nesting and Winograd nesting of small Winograd transforms. In step (ii), they further decomposed the pre-addition matrix **A** and post-addition matrix **C** into several factors, so that the number of design options available becomes very large: the N-point DFT when N has four factors can be calculated in over 10^{12} distinct ways.

This large family of nested algorithms contains the prime factor algorithm and the Winograd algorithms as particular cases, but usually achieves greater efficiency than either by reducing the f.p. multiplication count while keeping the number of f.p. additions small.

There is little doubt that this systematic approach will be extended so as to incorporate all available methods of restructuring the DFT.

1.3.4. Crystallographic applications of Fourier transforms

1.3.4.1. Introduction

The central role of the Fourier transformation in X-ray crystallography is a consequence of the kinematic approximation used in the description of the scattering of X-rays by a distribution of electrons (Bragg, 1915; Duane, 1925; Havighurst, 1925*a*,*b*; Zachariasen, 1945; James, 1948*a*, Chapters 1 and 2; Lipson & Cochran, 1953, Chapter 1; Bragg, 1975).

Let $\rho(\mathbf{X})$ be the density of electrons in a sample of matter contained in a finite region V which is being illuminated by a parallel monochromatic X-ray beam with wavevector $\mathbf{K}_0$. Then the far-field amplitude scattered in a direction corresponding to wavevector $\mathbf{K} = \mathbf{K}_0 + \mathbf{H}$ is proportional to

$$\begin{aligned} F(\mathbf{H}) &= \int_V \rho(\mathbf{X}) \exp(2\pi i \mathbf{H}\cdot\mathbf{X})\, \mathrm{d}^3\mathbf{X} \\ &= \bar{\mathscr{F}}[\rho](\mathbf{H}) \\ &= \langle \rho_{\mathbf{x}}, \exp(2\pi i \mathbf{H}\cdot\mathbf{X})\rangle. \end{aligned}$$

In certain model calculations, the 'sample' may contain not only volume charges, but also point, line and surface charges. These singularities may be accommodated by letting ρ be a distribution, and writing

$$F(\mathbf{H}) = \bar{\mathscr{F}}[\rho](\mathbf{H}) = \langle \rho_{\mathbf{x}}, \exp(2\pi i \mathbf{H}\cdot\mathbf{X})\rangle.$$

F is still a well behaved function (analytic, by Section 1.3.2.4.2.10) because ρ has been assumed to have compact support.

If the sample is assumed to be an infinite crystal, so that ρ is now a *periodic* distribution, the customary limiting process by which it is shown that F becomes a discrete series of peaks at reciprocal-lattice points (see *e.g.* von Laue, 1936; Ewald, 1940; James, 1948*a* p. 9; Lipson & Taylor, 1958, pp. 14–27; Ewald, 1962, pp. 82–101; Warren, 1969, pp. 27–30) is already subsumed under the treatment of Section 1.3.2.6.

1.3.4.2. Crystallographic Fourier transform theory

1.3.4.2.1. Crystal periodicity

1.3.4.2.1.1. Period lattice, reciprocal lattice and structure factors

Let ρ be the distribution of electrons in a crystal. Then, by definition of a crystal, ρ is Λ-*periodic* for some period lattice Λ (Section 1.3.2.6.5) so that there exists a motif distribution ρ^0 with compact support such that

$$\rho = R * \rho^0,$$

where $R = \sum_{\mathbf{x}\in\Lambda} \delta_{(\mathbf{X})}$. The lattice Λ is usually taken to be the finest for which the above representation holds.

Let Λ have a basis $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ over the integers, these basis vectors being expressed in terms of a standard orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ as

$$\mathbf{a}_k = \sum_{j=1}^{3} a_{jk}\mathbf{e}_j.$$

Then the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

is the period matrix of Λ (Section 1.3.2.6.5) with respect to the unit lattice with basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, and the volume V of the unit cell is given by $V = |\det \mathbf{A}|$.

By Fourier transformation

$$\bar{\mathscr{F}}[\rho] = R^* \times \bar{\mathscr{F}}[\rho^0],$$

where $R^* = \sum_{\mathbf{H}\in\Lambda^*} \delta_{(\mathbf{H})}$ is the lattice distribution associated to the reciprocal lattice Λ^*. The basis vectors $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ have coordinates in $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ given by the columns of $(\mathbf{A}^{-1})^T$, whose expression in terms of the cofactors of **A** (see Section 1.3.2.6.5) gives the familiar formulae involving the cross product of vectors for $n = 3$. The **H**-distribution F of scattered amplitudes may be written

$$F = \bar{\mathscr{F}}[\rho]_{\mathbf{H}} = \sum_{\mathbf{H}\in\Lambda^*} \bar{\mathscr{F}}[\rho^0](\mathbf{H})\delta_{(\mathbf{H})} = \sum_{\mathbf{H}\in\Lambda^*} F_{\mathbf{H}}\delta_{(\mathbf{H})}$$

and is thus a weighted reciprocal-lattice distribution, the weight $F_{\mathbf{H}}$ attached to each node $\mathbf{H} \in \Lambda^*$ being the value at **H** of the transform $\bar{\mathscr{F}}[\rho^0]$ of the motif ρ^0. Taken in conjunction with the assumption that the scattering is elastic, *i.e.* that **H** only changes the direction but not the magnitude of the incident wavevector $\mathbf{K}_0$, this result yields the usual forms (Laue or Bragg) of the

diffraction conditions: $\mathbf{H} \in \Lambda^*$, and simultaneously $\mathbf{H}$ lies on the Ewald sphere.

By the reciprocity theorem, ρ^0 can be recovered if F is known for all $\mathbf{H} \in \Lambda^*$ as follows [Section 1.3.2.6.5, *e.g.* (iv)]:

$$\rho_{\mathbf{x}} = \frac{1}{V} \sum_{\mathbf{H} \in \Lambda^*} F_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{X}).$$

These relations may be rewritten in terms of standard, or 'fractional crystallographic', coordinates by putting

$$\mathbf{X} = \mathbf{A}\mathbf{x}, \quad \mathbf{H} = (\mathbf{A}^{-1})^T \mathbf{h},$$

so that a unit cell of the crystal corresponds to $\mathbf{x} \in \mathbb{R}^3/\mathbb{Z}^3$, and that $\mathbf{h} \in \mathbb{Z}^3$. Defining $\mathscr{p}$ and $\mathscr{p}^0$ by

$$\rho = \frac{1}{V} A^{\#} \mathscr{p}, \quad \rho^0 = \frac{1}{V} A^{\#} \mathscr{p}^0$$

so that

$$\rho(\mathbf{X})\, \mathrm{d}^3\mathbf{X} = \mathscr{p}(\mathbf{x})\, \mathrm{d}^3\mathbf{x}, \quad \rho^0(\mathbf{X})\, \mathrm{d}^3\mathbf{X} = \mathscr{p}^0(\mathbf{x})\, \mathrm{d}^3\mathbf{x},$$

we have

$$\begin{aligned} \bar{\mathscr{F}}[\mathscr{p}]_{\mathbf{h}} &= \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \delta_{(\mathbf{h})}, \\ F(\mathbf{h}) &= \langle \mathscr{p}^0_{\mathbf{x}}, \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) \rangle \\ &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \mathscr{p}^0(\mathbf{x}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x})\, \mathrm{d}^3\mathbf{x} \quad \text{if } \mathscr{p}^0 \in L^1_{\mathrm{loc}}(\mathbb{R}^3/\mathbb{Z}^3), \\ \mathscr{p}_{\mathbf{x}} &= \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}). \end{aligned}$$

These formulae are valid for an arbitrary motif distribution $\mathscr{p}^0$, provided the convergence of the Fourier series for $\mathscr{p}$ is considered from the viewpoint of distribution theory (Section 1.3.2.6.10.3).

The experienced crystallographer may notice the absence of the familiar factor $1/V$ from the expression for $\mathscr{p}$ just given. This is because we use the (mathematically) natural unit for $\mathscr{p}$, the *electron per unit cell*, which matches the dimensionless nature of the crystallographic coordinates $\mathbf{x}$ and of the associated volume element $\mathrm{d}^3\mathbf{x}$. The traditional factor $1/V$ was the result of the somewhat inconsistent use of $\mathbf{x}$ as an argument but of $\mathrm{d}^3\mathbf{X}$ as a volume element to obtain ρ in electrons per unit volume (*e.g.* Å^3). A fortunate consequence of the present convention is that nuisance factors of V or $1/V$, which used to abound in convolution or scalar product formulae, are now absent.

It should be noted at this point that the crystallographic terminology regarding $\mathscr{F}$ and $\bar{\mathscr{F}}$ differs from the standard mathematical terminology introduced in Section 1.3.2.4.1 and applied to periodic distributions in Section 1.3.2.6.4: F is the *inverse* Fourier transform of ρ rather than its Fourier transform, and the calculation of ρ is called a Fourier *synthesis* in crystallography even though it is mathematically a Fourier analysis. The origin of this discrepancy may be traced to the fact that the mathematical theory of the Fourier transformation originated with the study of *temporal* periodicity, while crystallography deals with *spatial* periodicity; since the expression for the phase factor of a plane wave is $\exp[2\pi i(\nu t - \mathbf{K} \cdot \mathbf{X})]$, the difference in sign between the contributions from time *versus* spatial displacements makes this conflict unavoidable.

1.3.4.2.1.2. Structure factors in terms of form factors

In many cases, $\mathscr{p}^0$ is a sum of translates of atomic electron-density distributions. Assume there are n distinct chemical types of atoms, with N_j identical isotropic atoms of type j described by an electron distribution $\mathscr{p}_j$ about their centre of mass. According to quantum mechanics each $\mathscr{p}_j$ is a smooth rapidly decreasing function of $\mathbf{x}$, *i.e.* $\mathscr{p}_j \in \mathscr{S}$, hence $\mathscr{p}^0 \in \mathscr{S}$ and (ignoring the effect of thermal agitation)

$$\mathscr{p}^0(\mathbf{x}) = \sum_{j=1}^{n} \left[\sum_{k_j=1}^{N_j} \mathscr{p}_j(\mathbf{x} - \mathbf{x}_{k_j}) \right],$$

which may be written (Section 1.3.2.5.8)

$$\mathscr{p}^0 = \sum_{j=1}^{n} \left[\mathscr{p}_j * \left(\sum_{k_j=1}^{N_j} \delta_{(\mathbf{x}_{k_j})} \right) \right].$$

By Fourier transformation:

$$F(\mathbf{h}) = \sum_{j=1}^{n} \left\{ \bar{\mathscr{F}}[\mathscr{p}_j](\mathbf{h}) \times \left[\sum_{k_j=1}^{N_j} \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_{k_j}) \right] \right\}.$$

Defining the form factor f_j of atom j as a function of $\mathbf{h}$ to be

$$f_j(\mathbf{h}) = \bar{\mathscr{F}}[\mathscr{p}_j](\mathbf{h})$$

we have

$$F(\mathbf{h}) = \sum_{j=1}^{n} f_j(\mathbf{h}) \times \left[\sum_{k_j=1}^{N_j} \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_{k_j}) \right].$$

If $\mathbf{X} = \mathbf{A}\mathbf{x}$ and $\mathbf{H} = (\mathbf{A}^{-1})^T \mathbf{h}$ are the real- and reciprocal-space coordinates in Å and Å^{-1}, and if $\rho_j(\|\mathbf{X}\|)$ is the spherically symmetric electron-density function for atom type j, then

$$f_j(\mathbf{H}) = \int_0^{\infty} 4\pi \|\mathbf{X}\|^2 \rho_j(\|\mathbf{X}\|) \frac{\sin(2\pi \|\mathbf{H}\| \|\mathbf{X}\|)}{2\pi \|\mathbf{H}\| \|\mathbf{X}\|}\, \mathrm{d}\|\mathbf{X}\|.$$

More complex expansions are used for electron-density studies (see Chapter 1.2 in this volume). Anisotropic Gaussian atoms may be dealt with through the formulae given in Section 1.3.2.4.4.2.

1.3.4.2.1.3. Fourier series for the electron density and its summation

The convergence of the Fourier series for $\mathscr{p}$

$$\mathscr{p}(\mathbf{x}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

is usually examined from the classical point of view (Section 1.3.2.6.10). The summation of multiple Fourier series meets with considerable difficulties, because there is no natural order in $\mathbb{Z}^n$ to play the role of the natural order in $\mathbb{Z}$ (Ash, 1976). In crystallography, however, the structure factors $F(\mathbf{h})$ are often obtained within spheres $\|\mathbf{H}\| \leq \Delta^{-1}$ for increasing resolution (decreasing Δ). Therefore, successive estimates of $\mathscr{p}$ are most

naturally calculated as the corresponding partial sums (Section 1.3.2.6.10.1):

$$S_\Delta(\rho)(\mathbf{x}) = \sum_{\|(\mathbf{A}^{-1})^T\mathbf{h}\| \le \Delta^{-1}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}).$$

This may be written

$$S_\Delta(\rho)(\mathbf{x}) = (D_\Delta * \rho)(\mathbf{x}),$$

where D_Δ is the 'spherical Dirichlet kernel'

$$D_\Delta(\mathbf{x}) = \sum_{\|(\mathbf{A}^{-1})^T\mathbf{h}\| \le \Delta^{-1}} \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}).$$

D_Δ exhibits numerous negative ripples around its central peak. Thus the 'series termination errors' incurred by using $S_\Delta(\rho)$ instead of ρ consist of negative ripples around each atom, and may lead to a Gibbs-like phenomenon (Section 1.3.2.6.10.1) near a molecular boundary.

As in one dimension, Cesàro sums (arithmetic means of partial sums) have better convergence properties, as they lead to a convolution by a 'spherical Fejér kernel' which is everywhere positive. Thus Cesàro summation will always produce positive approximations to a positive electron density. Other positive summation kernels were investigated by Pepinsky (1952) and by Waser & Schomaker (1953).

1.3.4.2.1.4. Friedel's law, anomalous scatterers

If the wavelength λ of the incident X-rays is far from any absorption edge of the atoms in the crystal, there is a constant phase shift in the scattering, and the electron density may be considered to be *real-valued*. Then

$$\begin{aligned} F(\mathbf{h}) &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \exp(2\pi i \mathbf{h}\cdot\mathbf{x})\, \mathrm{d}^3\mathbf{x} \\ &= \overline{\int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho(\mathbf{x})} \exp[2\pi i(-\mathbf{h})\cdot\mathbf{x}]\, \mathrm{d}^3\mathbf{x}} \\ &= \overline{F(-\mathbf{h})} \text{ since } \overline{\rho(\mathbf{x})} = \rho(\mathbf{x}). \end{aligned}$$

Thus if

$$F(\mathbf{h}) = |F(\mathbf{h})| \exp(i\varphi(\mathbf{h})),$$

then

$$|F(-\mathbf{h})| = |F(\mathbf{h})| \quad \text{and} \quad \varphi(-\mathbf{h}) = -\varphi(\mathbf{h}).$$

This is Friedel's law (Friedel, 1913). The set $\{F_\mathbf{h}\}$ of Fourier coefficients is said to have *Hermitian symmetry*.

If λ is close to some absorption edge(s), the proximity to resonance induces an extra phase shift, whose effect may be represented by letting $\rho(\mathbf{x})$ take on *complex values*. Let

$$\rho(\mathbf{x}) = \rho^{\mathrm{R}}(\mathbf{x}) + i\rho^{\mathrm{I}}(\mathbf{x})$$

and correspondingly, by termwise Fourier transformation

$$F(\mathbf{h}) = F^{\mathrm{R}}(\mathbf{h}) + iF^{\mathrm{I}}(\mathbf{h}).$$

Since $\rho^{\mathrm{R}}(\mathbf{x})$ and $\rho^{\mathrm{I}}(\mathbf{x})$ are both real, $F^{\mathrm{R}}(\mathbf{h})$ and $F^{\mathrm{I}}(\mathbf{h})$ are both Hermitian symmetric, hence

$$F(-\mathbf{h}) = \overline{F^{\mathrm{R}}(\mathbf{h})} + \overline{iF^{\mathrm{I}}(\mathbf{h})},$$

while

$$\overline{F(\mathbf{h})} = \overline{F^{\mathrm{R}}(\mathbf{h})} - \overline{iF^{\mathrm{I}}(\mathbf{h})}.$$

Thus $F(-\mathbf{h}) \neq \overline{F(\mathbf{h})}$, so that Friedel's law is violated. The components $F^{\mathrm{R}}(\mathbf{h})$ and $F^{\mathrm{I}}(\mathbf{h})$, which do obey Friedel's law, may be expressed as:

$$\begin{aligned} F^{\mathrm{R}}(\mathbf{h}) &= \tfrac{1}{2}[F(\mathbf{h}) + \overline{F(-\mathbf{h})}], \\ F^{\mathrm{I}}(\mathbf{h}) &= \frac{1}{2i}[F(\mathbf{h}) - \overline{F(-\mathbf{h})}]. \end{aligned}$$

1.3.4.2.1.5. Parseval's identity and other L^2 theorems

By Section 1.3.2.4.3.3 and Section 1.3.2.6.10.2,

$$\sum_{\mathbf{h}\in\mathbb{Z}^3} |F(\mathbf{h})|^2 = \int_{\mathbb{R}^3/\mathbb{Z}^3} |\rho(\mathbf{x})|^2\, \mathrm{d}^3\mathbf{x} = V \int_{\mathbb{R}^3/\Lambda} |\rho(\mathbf{X})|^2\, \mathrm{d}^3\mathbf{X}.$$

Usually $\rho(\mathbf{x})$ is real and positive, hence $|\rho(\mathbf{x})| = \rho(\mathbf{x})$, but the identity remains valid even when $\rho(\mathbf{x})$ is made complex-valued by the presence of anomalous scatterers.

If $\{G_\mathbf{h}\}$ is the collection of structure factors belonging to another electron density $\sigma = A^{\#}\sigma$ with the same period lattice as ρ, then

$$\begin{aligned} \sum_{\mathbf{h}\in\mathbb{Z}^3} \overline{F(\mathbf{h})} G(\mathbf{h}) &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho(\mathbf{x})}\sigma(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \\ &= V \int_{\mathbb{R}^3/\Lambda} \rho(\mathbf{X})\sigma(\mathbf{X})\, \mathrm{d}^3\mathbf{X}. \end{aligned}$$

Thus, norms and inner products may be evaluated either from structure factors or from 'maps'.

1.3.4.2.1.6. Convolution, correlation and Patterson function

Let $\rho = r * \rho^0$ and $\sigma = r * \sigma^0$ be two electron densities referred to crystallographic coordinates, with structure factors $\{F_\mathbf{h}\}_{\mathbf{h}\in\mathbb{Z}^3}$ and $\{G_\mathbf{h}\}_{\mathbf{h}\in\mathbb{Z}^3}$, so that

$$\begin{aligned} \rho_\mathbf{x} &= \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}), \\ \sigma_\mathbf{x} &= \sum_{\mathbf{h}\in\mathbb{Z}^3} G(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}). \end{aligned}$$

The distribution $\omega = r * (\rho^0 * \sigma^0)$ is well defined, since the generalized support condition (Section 1.3.2.3.9.7) is satisfied. The forward version of the convolution theorem implies that if

$$\omega_\mathbf{x} = \sum_{\mathbf{h}\in\mathbb{Z}^3} W(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}),$$

then

$$W(\mathbf{h}) = F(\mathbf{h})G(\mathbf{h}).$$

If either $\mathscr{p}^0$ or $\mathscr{o}^0$ is infinitely differentiable, then the distribution $\psi = \mathscr{p} \times \mathscr{o}$ exists, and if we analyse it as

$$\psi_{\mathbf{x}} = \sum_{\mathbf{h}\in\mathbb{Z}^3} Y(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}),$$

then the backward version of the convolution theorem reads:

$$Y(\mathbf{h}) = \sum_{\mathbf{k}\in\mathbb{Z}^3} F(\mathbf{h})G(\mathbf{h}-\mathbf{k}).$$

The *cross correlation* $\kappa[\mathscr{p},\mathscr{o}]$ between $\mathscr{p}$ and $\mathscr{o}$ is the $\mathbb{Z}^3$-periodic distribution defined by:

$$\kappa = \check{\mathscr{p}}^0 * \mathscr{o}.$$

If $\mathscr{p}^0$ and $\mathscr{o}^0$ are locally integrable,

$$\begin{aligned}\kappa[\mathscr{p},\mathscr{o}](\mathbf{t}) &= \int_{\mathbb{R}^3} \mathscr{p}^0(\mathbf{x})\mathscr{o}(\mathbf{x}+\mathbf{t})\,\mathrm{d}^3\mathbf{x} \\ &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \mathscr{p}(\mathbf{x})\mathscr{o}(\mathbf{x}+\mathbf{t})\,\mathrm{d}^3\mathbf{x}.\end{aligned}$$

Let

$$\kappa(\mathbf{t}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} K(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{t}).$$

The combined use of the shift property and of the forward convolution theorem then gives immediately:

$$K(\mathbf{h}) = \overline{F(\mathbf{h})}G(\mathbf{h});$$

hence the Fourier series representation of $\kappa[\mathscr{p},\mathscr{o}]$:

$$\kappa[\mathscr{p},\mathscr{o}](\mathbf{t}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} \overline{F(\mathbf{h})}G(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{t}).$$

Clearly, $\kappa[\mathscr{p},\mathscr{o}] = (\kappa[\mathscr{o},\mathscr{p}])\check{}$, as shown by the fact that permuting F and G changes $K(\mathbf{h})$ into its complex conjugate.

The *auto-correlation* of $\mathscr{p}$ is defined as $\kappa[\mathscr{p},\mathscr{p}]$ and is called the *Patterson function* of $\mathscr{p}$. If $\mathscr{p}$ consists of point atoms, *i.e.*

$$\mathscr{p}^0 = \sum_{j=1}^{N} Z_j \delta_{(\mathbf{x}_j)},$$

then

$$\kappa[\mathscr{p},\mathscr{p}] = r * \left[\sum_{j=1}^{N}\sum_{k=1}^{N} Z_j Z_k \delta_{(\mathbf{x}_j-\mathbf{x}_k)}\right]$$

contains information about interatomic vectors. It has the Fourier series representation

$$\kappa[\mathscr{p},\mathscr{p}](\mathbf{t}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{t}),$$

and is therefore calculable from the diffraction intensities alone. It was first proposed by Patterson (1934, 1935*a*,*b*) as an extension to crystals of the radially averaged correlation function used by Warren & Gingrich (1934) in the study of powders.

1.3.4.2.1.7. Sampling theorems, continuous transforms, interpolation

Shannon's sampling and interpolation theorem (Section 1.3.2.7.1) takes two different forms, according to whether the property of finite bandwidth is assumed in real space or in reciprocal space.

(1) The most usual setting is in reciprocal space (see Sayre, 1952*c*). Only a finite number of diffraction intensities can be recorded and phased, and for physical reasons the cutoff criterion is the resolution $\Delta = 1/\|\mathbf{H}\|_{\max}$. Electron-density maps are thus calculated as partial sums (Section 1.3.4.2.1.3), which may be written in Cartesian coordinates as

$$S_\Delta(\rho)(\mathbf{X}) = \sum_{\mathbf{H}\in\Lambda^*,\ \|\mathbf{H}\|\le\Delta^{-1}} F(\mathbf{H}) \exp(-2\pi i \mathbf{H}\cdot\mathbf{X}).$$

$S_\Delta(\rho)$ is band-limited, the support of its spectrum being contained in the solid sphere Σ_Δ defined by $\|\mathbf{H}\| \le \Delta^{-1}$. Let χ_Δ be the indicator function of Σ_Δ. The transform of the normalized version of χ_Δ is (see below, Section 1.3.4.4.3.5)

$$\begin{aligned}I_\Delta(\mathbf{X}) &= \frac{3\Delta^3}{4\pi}\mathscr{F}[\chi_\Delta](\mathbf{X}) \\ &= \frac{3}{u^3}(\sin u - u\cos u) \quad \text{where } u = 2\pi\frac{\|\mathbf{X}\|}{\Delta}.\end{aligned}$$

By Shannon's theorem, it suffices to calculate $S_\Delta(\rho)$ on an integral subdivision Γ of the period lattice Λ such that the sampling criterion is satisfied (*i.e.* that the translates of Σ_Δ by vectors of Γ^* do not overlap). Values of $S_\Delta(\rho)$ may then be calculated at an arbitrary point $\mathbf{X}$ by the interpolation formula:

$$S_\Delta(\rho)(\mathbf{X}) = \sum_{\mathbf{Y}\in\Gamma} I_\Delta(\mathbf{X}-\mathbf{Y}) S_\Delta(\rho)(\mathbf{Y}).$$

(2) The reverse situation occurs whenever the support of the motif $\mathscr{p}^0$ does not fill the whole unit cell, *i.e.* whenever there exists a region M (the 'molecular envelope'), strictly smaller than the unit cell, such that the translates of M by vectors of r do not overlap and that

$$\chi_M \times \mathscr{p}^0 = \mathscr{p}^0.$$

It then follows that $\mathscr{p} = r * (\chi_M \times \mathscr{p})$. Defining the 'interference function' G as the normalized indicator function of M according to

$$G(\boldsymbol{\eta}) = \frac{1}{\mathrm{vol}(M)}\bar{\mathscr{F}}[\chi_M](\boldsymbol{\eta})$$

we may invoke Shannon's theorem to calculate the value $\bar{\mathscr{F}}[\mathscr{p}^0](\boldsymbol{\xi})$ at an arbitrary point $\boldsymbol{\xi}$ of reciprocal space from its sample values $F(\mathbf{h}) = \bar{\mathscr{F}}[\mathscr{p}^0](\mathbf{h})$ at points of the reciprocal lattice as

$$\bar{\mathscr{F}}[\mathscr{p}^0](\boldsymbol{\xi}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} \mathbf{G}(\boldsymbol{\xi}-\mathbf{h})F(\mathbf{h}).$$

This aspect of Shannon's theorem constitutes the mathematical basis of phasing methods based on geometric redundancies created by solvent regions and/or noncrystallographic symmetries (Bricogne, 1974). The connection between Shannon's theorem and the phase problem was first noticed by Sayre (1952*b*). He pointed out that the Patterson function of ρ, written as $\kappa[\rho, \rho] = r * (\check{\rho}^0 * \rho^0)$, may be viewed as consisting of a motif $\kappa^0 = \check{\rho}^0 * \rho^0$ (containing all the internal interatomic vectors) which is periodized by convolution with r. As the translates of κ^0 by vectors of $\mathbb{Z}^3$ *do* overlap, the sample values of the intensities $|F(\mathbf{h})|^2$ at nodes of the reciprocal lattice do not provide enough data to interpolate intensities $|F(\boldsymbol{\xi})|^2$ at arbitrary points of reciprocal space. Thus the loss of phase is intimately related to the impossibility of intensity interpolation, implying in return that any indication of intensity values attached to nonintegral points of the reciprocal lattice is a potential source of phase information.

1.3.4.2.1.8. Sections and projections

It was shown at the end of Section 1.3.2.5.8 that the convolution theorem establishes, under appropriate assumptions, a duality between sectioning a smooth function (viewed as a multiplication by a δ-function in the sectioning coordinate) and projecting its transform (viewed as a convolution with the function **1** everywhere equal to 1 as a function of the projection coordinate). This duality follows from the fact that $\mathcal{F}$ and $\bar{\mathcal{F}}$ map $\mathbf{1}_{x_i}$ to δ_{x_i} and δ_{x_i} to $\mathbf{1}_{x_i}$ (Section 1.3.2.5.6), and from the tensor product property (Section 1.3.2.5.5).

In the case of periodic distributions, projection and section must be performed with respect to directions or subspaces which are integral with respect to the period lattice if the result is to be periodic; furthermore, projections must be performed only on the contents of *one* repeating unit along the direction of projection, or else the result would diverge. The same relations then hold between principal central sections and projections of the electron density and the dual principal central projections and sections of the weighted reciprocal lattice, *e.g.*

$$\begin{aligned}
\rho(x_1, 0, 0) &\leftrightarrow \sum_{h_1, h_2} F(h_1, h_2, h_3), \\
\rho(x_1, x_2, 0) &\leftrightarrow \sum_{h_3} F(h_1, h_2, h_3), \\
\rho_{1,2}(x_3) = \int_{\mathbb{R}^2/\mathbb{Z}^2} \rho(x_1, x_2, x_3)\, \mathrm{d}x_1\, \mathrm{d}x_2 &\leftrightarrow F(0, 0, h_3), \\
\rho_1(x_2, x_3) = \int_{\mathbb{R}/\mathbb{Z}} \rho(x_1, x_2, x_3)\, \mathrm{d}x_1 &\leftrightarrow F(0, h_2, h_3)
\end{aligned}$$

etc.

When the sections are principal but not central, it suffices to use the shift property of Section 1.3.2.5.5. When the sections or projections are not principal, they can be made principal by changing to new primitive bases B and B^* for Λ and Λ^*, respectively, the transition matrices $\mathbf{P}$ and $\mathbf{P}^*$ to these new bases being related by $\mathbf{P}^* = (\mathbf{P}^{-1})^T$ in order to preserve duality. This change of basis must be such that one of these matrices (say, $\mathbf{P}$) should have a given integer vector $\mathbf{u}$ as its first column, $\mathbf{u}$ being related to the line or plane defining the section or projection of interest.

The problem of constructing a matrix $\mathbf{P}$ given $\mathbf{u}$ received an erroneous solution in Volume II of *International Tables* (Patterson, 1959), which was subsequently corrected in 1962. Unfortunately, the solution proposed there is complicated and does not suggest a general approach to the problem. It therefore seems worthwhile to record here an effective procedure which solves this problem in any dimension n (Watson, 1970).

Let

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

be a primitive integral vector, *i.e.* g.c.d. $(u_1, \ldots, u_n) = 1$. Then an $n \times n$ integral matrix $\mathbf{P}$ with $\det \mathbf{P} = 1$ having $\mathbf{u}$ as its first column can be constructed by induction as follows. For $n = 1$ the result is trivial. For $n = 2$ it can be solved by means of the Euclidean algorithm, which yields z_1, z_2 such that $u_1 z_2 - u_2 z_1 = 1$, so that we may take

$$\mathbf{P} = \begin{pmatrix} u_1 & z_1 \\ u_2 & z_2 \end{pmatrix}.$$

Note that, if

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

is a solution, then $\mathbf{z} + m\mathbf{u}$ is another solution for any $m \in \mathbb{Z}$. For $n \geq 3$, write

$$\mathbf{u} = \begin{pmatrix} u_1 \\ d\mathbf{z} \end{pmatrix}$$

with $d =$ g.c.d. $(u_2, \ldots, u_n)$ so that both

$$\mathbf{z} = \begin{pmatrix} z_2 \\ \vdots \\ z_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u_1 \\ d \end{pmatrix}$$

are primitive. By the inductive hypothesis there is an integral 2×2 matrix $\mathbf{V}$ with

$$\begin{pmatrix} u_1 \\ d \end{pmatrix}$$

as its first column, and an integral $(n-1) \times (n-1)$ matrix $\mathbf{Z}$ with $\mathbf{z}$ as its first column, with $\det \mathbf{V} = 1$ and $\det \mathbf{Z} = 1$.

Now put

$$\mathbf{P} = \begin{pmatrix} 1 & \\ & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{V} & \\ & \mathbf{I}_{n-2} \end{pmatrix},$$

i.e.

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & . & 0 \\ 0 & z_2 & * & . & * \\ 0 & z_3 & * & . & * \\ . & . & . & . & . \\ 0 & z_n & * & . & * \end{pmatrix} \begin{pmatrix} u_1 & * & 0 & . & 0 \\ d & * & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & . & 1 \end{pmatrix}.$$

The first column of $\mathbf{P}$ is

$$\begin{pmatrix} u_1 \\ dz_2 \\ . \\ . \\ dz_n \end{pmatrix} = \mathbf{u},$$

and its determinant is 1, QED.

The incremental step from dimension $n-1$ to dimension n is the construction of 2×2 matrix $\mathbf{V}$, for which there exist infinitely many solutions labelled by an integer m_{n-1}. Therefore, the collection of matrices $\mathbf{P}$ which solve the problem is labelled by $n-1$ arbitrary integers $(m_1, m_2, \ldots, m_{n-1})$. This freedom can be used to adjust the shape of the basis B.

Once $\mathbf{P}$ has been chosen, the calculation of general sections and projections is transformed into that of *principal* sections and projections by the changes of coordinates:

$$\mathbf{x} = \mathbf{P}\mathbf{x}', \qquad \mathbf{h} = \mathbf{P}^*\mathbf{h}',$$

and an appeal to the tensor product property.

Booth (1945*a*) made use of the convolution theorem to form the Fourier coefficients of 'bounded projections', which provided a compromise between 2D and 3D Fourier syntheses. If it is desired to compute the projection on the (x, y) plane of the electron density lying between the planes $z = z_1$ and $z = z_2$, which may be written as

$$[\rho \times (\mathbf{1}_x \otimes \mathbf{1}_y \otimes \chi_{[z_1, z_2]})] * (\delta_x \otimes \delta_y \otimes \mathbf{1}_z).$$

The transform is then

$$[F * (\delta_h \otimes \delta_k \otimes \bar{\mathscr{F}}[\chi_{[z_1, z_2]}])] \times (\mathbf{1}_h \otimes \mathbf{1}_k \otimes \delta_l),$$

giving for coefficient (h, k):

$$\sum_{l \in \mathbb{Z}} F(h, k, l) \exp\{2\pi i l[(z_1 + z_2)/2]\} \times \frac{\sin \pi l(z_1 - z_2)}{\pi l}.$$

1.3.4.2.1.9. Differential syntheses

Another particular instance of the convolution theorem is the duality between differentiation and multiplication by a monomial (Sections 1.3.2.4.2.8, 1.3.2.5.8).

In the present context, this result may be written

$$\bar{\mathscr{F}}\left[\frac{\partial^{m_1+m_2+m_3}\rho}{\partial X_1^{m_1}\partial X_2^{m_2}\partial X_3^{m_3}}\right](\mathbf{H}) = (-2\pi i)^{m_1+m_2+m_3} H_1^{m_1} H_2^{m_2} H_3^{m_3} F(\mathbf{A}^T\mathbf{H})$$

in Cartesian coordinates, and

$$\bar{\mathscr{F}}\left[\frac{\partial^{m_1+m_2+m_3}\rho}{\partial x_1^{m_1}\partial x_2^{m_2}\partial x_3^{m_3}}\right](\mathbf{h}) = (-2\pi i)^{m_1+m_2+m_3} h_1^{m_1} h_2^{m_2} h_3^{m_3} F(\mathbf{h})$$

in crystallographic coordinates.

A particular case of the first formula is

$$-4\pi^2 \sum_{\mathbf{H} \in \Lambda^*} \|\mathbf{H}\|^2 F(\mathbf{A}^T\mathbf{H}) \exp(-2\pi i \mathbf{H} \cdot \mathbf{X}) = \Delta\rho(\mathbf{X}),$$

where

$$\Delta\rho = \sum_{j=1}^{3} \frac{\partial^2 \rho}{\partial X_j^2}$$

is the Laplacian of ρ.

The second formula has been used with $|\mathbf{m}| = 1$ or 2 to compute 'differential syntheses' and refine the location of maxima (or other stationary points) in electron-density maps. Indeed, the values at $\mathbf{x}$ of the gradient vector $\nabla\rho$ and Hessian matrix $(\nabla\nabla^T)\rho$ are readily obtained as

$$(\nabla\rho)(\mathbf{x}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} (-2\pi i \mathbf{h}) F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}),$$
$$[(\nabla\nabla^T)\rho](\mathbf{x}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} (-4\pi^2 \mathbf{h}\mathbf{h}^T) F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}),$$

and a step of Newton iteration towards the nearest stationary point of ρ will proceed by

$$\mathbf{x} \longmapsto \mathbf{x} - \{[(\nabla\nabla^T)\rho](\mathbf{x})\}^{-1}(\nabla\rho)(\mathbf{x}).$$

The modern use of Fourier transforms to speed up the computation of derivatives for model refinement will be described in Section 1.3.4.4.7.

The converse property is also useful: it relates the derivatives of the continuous transform $\bar{\mathscr{F}}[\rho^0]$ to the moments of ρ^0:

$$\frac{\partial^{m_1+m_2+m_3}\bar{\mathscr{F}}[\rho^0]}{\partial X_1^{m_1}\partial X_2^{m_2}\partial X_3^{m_3}}(\mathbf{H}) = \bar{\mathscr{F}}[(2\pi i)^{m_1+m_2+m_3} X_1^{m_1} X_2^{m_2} X_3^{m_3} \rho_{\mathbf{x}}^0](\mathbf{H}).$$

For $|\mathbf{m}| = 2$ and $\mathbf{H} = \mathbf{0}$, this identity gives the well known relation between the Hessian matrix of the transform $\bar{\mathscr{F}}[\rho^0]$ at the origin of reciprocal space and the inertia tensor of the motif ρ^0. This is a particular case of the moment-generating properties of $\bar{\mathscr{F}}$, which will be further developed in Section 1.3.4.5.2.

1.3.4.2.1.10. Toeplitz forms, determinantal inequalities and Szegö's theorem

The classical results presented in Section 1.3.2.6.9 can be readily generalized to the case of triple Fourier series; no new concept is needed, only an obvious extension of the notation.

Let ρ be real-valued, so that Friedel's law holds and $F(-\mathbf{h}) = \overline{F(\mathbf{h})}$. Let H be a finite set of indices comprising the origin: $\mathsf{H} = \{\mathbf{h}_0 = \mathbf{0}, \mathbf{h}_1, \ldots, \mathbf{h}_n\}$. Then the Hermitian form in $n+1$ complex variables

$$T_{\mathsf{H}}[\rho](\mathbf{u}) = \sum_{j,\,k=0}^{n} F(\mathbf{h}_j - \mathbf{h}_k)\overline{u_j}u_k$$

is called the Toeplitz form of order H associated to ρ. By the convolution theorem and Parseval's identity,

$$T_{\mathsf{H}}[\rho](\mathbf{u}) = \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \left| \sum_{j=0}^{n} u_j \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) \right|^2 \mathrm{d}^3\mathbf{x}.$$

If ρ is almost everywhere non-negative, then for all H the forms $T_{\mathsf{H}}[\rho]$ are positive semi-definite and therefore all Toeplitz determinants $D_{\mathsf{H}}[\rho]$ are non-negative, where

$$D_{\mathsf{H}}[\mathscr{P}] = \det\{[F(\mathbf{h}_j - \mathbf{h}_k)]\}.$$

The Toeplitz–Carathéodory–Herglotz theorem given in Section 1.3.2.6.9.2 states that the converse is true: if $D_{\mathsf{H}}[\rho] \geq 0$ for all H, then $\mathscr{P}$ is almost everywhere non-negative. This result is known in the crystallographic literature through the papers of Karle & Hauptman (1950), MacGillavry (1950), and Goedkoop (1950), following previous work by Harker & Kasper (1948) and Gillis (1948*a*,*b*).

Szegö's study of the asymptotic distribution of the eigenvalues of Toeplitz forms as their order tends to infinity remains valid. Some precautions are needed, however, to define the notion of a sequence (H_k) of finite subsets of indices tending to infinity: it suffices that the H_k should consist essentially of the reciprocal-lattice points $\mathbf{h}$ contained within a domain of the form $k\Omega$ (k-fold dilation of Ω) where Ω is a convex domain in $\mathbb{R}^3$ containing the origin (Widom, 1960). Under these circumstances, the eigenvalues $\lambda_\nu^{(n)}$ of the Toeplitz forms $T_{\mathsf{H}_k}[\mathscr{P}]$ become equidistributed with the sample values $\mathscr{P}_\nu^{(n)}$ of $\mathscr{P}$ on a grid satisfying the Shannon sampling criterion for the data in H_k (*cf.* Section 1.3.2.6.9.3).

A particular consequence of this equidistribution is that the geometric means of the $\lambda_\nu^{(n)}$ and of the $\mathscr{P}_\nu^{(n)}$ are equal, and hence as in Section 1.3.2.6.9.4

$$\lim_{k\to\infty}\{D_{\mathsf{H}_k}[\mathscr{P}]\}^{1/|\mathsf{H}_k|} = \exp\left\{\int_{\mathbb{R}^3/\mathbb{Z}^3} \log \mathscr{P}(\mathbf{x})\, \mathrm{d}^3\mathbf{x}\right\},$$

where $|\mathsf{H}_k|$ denotes the number of reflections in H_k. Complementary terms giving a better comparison of the two sides were obtained by Widom (1960, 1975) and Linnik (1975).

This formula played an important role in the solution of the 2D Ising model by Onsager (1944) (see Montroll *et al.*, 1963). It is also encountered in phasing methods involving the 'Burg entropy' (Britten & Collins, 1982; Narayan & Nityananda, 1982; Bricogne, 1982, 1984, 1988).

1.3.4.2.2. Crystal symmetry

1.3.4.2.2.1. Crystallographic groups

The description of a crystal given so far has dealt only with its invariance under the action of the (discrete Abelian) group of translations by vectors of its period lattice Λ.

Let the crystal now be embedded in Euclidean 3-space, so that it may be acted upon by the group $M(3)$ of rigid (*i.e.* distance-preserving) motions of that space. The group $M(3)$ contains a normal subgroup $T(3)$ of translations, and the quotient group $M(3)/T(3)$ may be identified with the 3-dimensional orthogonal group $O(3)$. The period lattice Λ of a crystal is a discrete uniform subgroup of $T(3)$.

The possible invariance properties of a crystal under the action of $M(3)$ are captured by the following definition: a *crystallographic group* is a subgroup Γ of $M(3)$ if

(i) $\Gamma \cap T(3) = \Lambda$, a period lattice and a normal subgroup of Γ;

(ii) the factor group $G = \Gamma/\Lambda$ is finite.

The two properties are not independent: by a theorem of Bieberbach (1911), they follow from the assumption that Λ is a discrete subgroup of $M(3)$ which operates without accumulation point and with a compact fundamental domain (see Auslander, 1965). These two assumptions imply that G acts on Λ through an integral representation, and this observation leads to a complete enumeration of all distinct Γ's. The mathematical theory of these groups is still an active research topic (see, for instance, Farkas, 1981), and has applications to Riemannian geometry (Wolf, 1967).

This classification of crystallographic groups is described elsewhere in these *Tables* (Wondratschek, 2005), but it will be surveyed briefly in Section 1.3.4.2.2.3 for the purpose of establishing further terminology and notation, after recalling basic notions and results concerning groups and group actions in Section 1.3.4.2.2.2.

1.3.4.2.2.2. Groups and group actions

The books by Hall (1959) and Scott (1964) are recommended as reference works on group theory.

(*a*) *Left and right actions*

Let G be a group with identity element e, and let X be a set. An *action* of G on X is a mapping from $G \times X$ to X with the property that, if $g\,x$ denotes the image of (g, x), then

(i) $(g_1g_2)x = g_1(g_2x)$ for all $g_1, g_2 \in G$ and all $x \in X$,

(ii) $ex = x$ for all $x \in X$.

An element g of G thus induces a mapping T_g of X into itself defined by $T_g(x) = gx$, with the 'representation property':

(iii) $T_{g_1g_2} = T_{g_1}T_{g_2}$ for all $g_1, g_2 \in G$.

Since G is a group, every g has an inverse g^{-1}; hence every mapping T_g has an inverse $T_{g^{-1}}$, so that each T_g is a permutation of X.

Strictly speaking, what has just been defined is a *left* action. A *right* action of G on X is defined similarly as a mapping $(g, x) \longmapsto xg$ such that

(i′) $x(g_1g_2) = (xg_1)g_2$ for all $g_1, g_2 \in G$ and all $x \in X$,

(ii′) $xe = x$ for all $x \in X$.

The mapping T'_g defined by $T'_g(x) = xg$ then has the 'right-representation' property:

(iii′) $T'_{g_1g_2} = T'_{g_2}T'_{g_1}$ for all $g_1, g_2 \in G$.

The essential difference between left and right actions is of course not whether the elements of G are written on the left or right of those of X: it lies in the difference between (iii) and (iii′). In a left action the product g_1g_2 in G operates on $x \in X$ by g_2 operating first, then g_1 operating on the result; in a right action, g_1 operates first, then g_2. This distinction will be of importance in Sections 1.3.4.2.2.4 and 1.3.4.2.2.5. In the sequel, we will use left actions unless otherwise stated.

(*b*) *Orbits and isotropy subgroups*

Let x be a fixed element of X. Two fundamental entities are associated to x:

(1) the subset of G consisting of all g such that $gx = x$ is a subgroup of G, called the *isotropy subgroup* of x and denoted G_x;

(2) the subset of X consisting of all elements gx with g running through G is called the *orbit* of x under G and is denoted Gx.

Through these definitions, the action of G on X can be related to the internal structure of G, as follows. Let G/G_x denote the collection of distinct left cosets of G_x in G, *i.e.* of distinct subsets of G of the form gG_x. Let $|G|$, $|G_x|$, $|Gx|$ and $|G/G_x|$ denote the numbers of elements in the corresponding sets. The number $|G/G_x|$ of distinct cosets of G_x in G is also denoted $[G : G_x]$ and is called the *index* of G_x in G; by Lagrange's theorem

$$[G : G_x] = |G/G_x| = \frac{|G|}{|G_x|}.$$

Now if g_1 and g_2 are in the same coset of G_x, then $g_2 = g_1 g'$ with $g' \in G_x$, and hence $g_1 x = g_2 x$; the converse is obviously true. Therefore, the mapping from cosets to orbit elements

$$gG_x \longmapsto gx$$

establishes a one-to-one correspondence between the distinct left cosets of G_x in G and the elements of the orbit of x under G. It follows that the number of distinct elements in the orbit of x is equal to the index of G_x in G:

$$|Gx| = [G : G_x] = \frac{|G|}{|G_x|},$$

and that the elements of the orbit of x may be listed without repetition in the form

$$Gx = \{\gamma x | \gamma \in G/G_x\}.$$

Similar definitions may be given for a right action of G on X. The set of distinct right cosets $G_x g$ in G, denoted $G_x\backslash G$, is then in one-to-one correspondence with the distinct elements in the orbit xG of x.

(*c*) *Fundamental domain and orbit decomposition*

The group properties of G imply that two orbits under G are either disjoint or equal. The set X may thus be written as the *disjoint* union

$$X = \bigcup_{i \in I} Gx_i,$$

where the x_i are elements of distinct orbits and I is an indexing set labelling them. The subset $D = \{x_i\}_{i \in I}$ is said to constitute a *fundamental domain* (mathematical terminology) or an *asymmetric unit* (crystallographic terminology) for the action of G on X: it contains one representative x_i of each distinct orbit. Clearly, other fundamental domains may be obtained by choosing different representatives for these orbits.

If X is finite and if f is an arbitrary complex-valued function over X, the 'integral' of f over X may be written as a sum of integrals over the distinct orbits, yielding the *orbit decomposition formula*:

$$\sum_{x \in X} f(x) = \sum_{i \in I} \left(\sum_{y_i \in Gx_i} f(y_i) \right) = \sum_{i \in I} \left(\sum_{\gamma_i \in G/G_{x_i}} f(\gamma_i x_i) \right)$$
$$= \sum_{i \in I} \frac{1}{|G_{x_i}|} \left(\sum_{g_i \in G} f(g_i x_i) \right).$$

In particular, taking $f(x) = 1$ for all x and denoting by $|X|$ the number of elements of X:

$$|X| = \sum_{i \in I} |Gx_i| = \sum_{i \in I} |G/G_{x_i}| = \sum_{i \in I} \frac{|G|}{|G_{x_i}|}.$$

(*d*) *Conjugation, normal subgroups, semi-direct products*

A group G acts on itself by *conjugation*, *i.e.* by associating to $g \in G$ the mapping C_g defined by

$$C_g(h) = ghg^{-1}.$$

Indeed, $C_g(hk) = C_g(h)C_g(k)$ and $[C_g(h)]^{-1} = C_{g^{-1}}(h)$. In particular, C_g operates on the set of subgroups of G, two subgroups H and K being called conjugate if $H = C_g(K)$ for some $g \in G$; for example, it is easily checked that $G_{gx} = C_g(G_x)$. The orbits under this action are the *conjugacy classes* of subgroups of G, and the isotropy subgroup of H under this action is called the *normalizer* of H in G.

If $\{H\}$ is a one-element orbit, H is called a *self-conjugate* or *normal* subgroup of G; the cosets of H in G then form a group G/H called the *factor group* of G by H.

Let G and H be two groups, and suppose that G acts on H by *automorphisms* of H, *i.e.* in such a way that

$$g(h_1 h_2) = g(h_1)g(h_2)$$
$$g(e_H) = e_H \qquad \text{(where } e_H \text{ is the identity element of } H\text{)}.$$
$$g(h^{-1}) = (g(h))^{-1}$$

Then the symbols $[g, h]$ with $g \in G$, $h \in H$ form a group K under the product rule:

$$[g_1, h_1][g_2, h_2] = [g_1 g_2, h_1 g_1(h_2)]$$

{associativity checks; $[e_G, e_H]$ is the identity; $[g, h]$ has inverse $[g^{-1}, g^{-1}(h^{-1})]$}. The group K is called the *semi-direct product* of H by G, denoted $K = H \rtimes G$.

The elements $[g, e_H]$ form a subgroup of K isomorphic to G, the elements $[e_G, h]$ form a normal subgroup of K isomorphic to H, and the action of G on H may be represented as an action by conjugation in the sense that

$$C_{[g, e_H]}([e_G, h]) = [e_G, g(h)].$$

A familiar example of semi-direct product is provided by the group of Euclidean motions $M(3)$ (Section 1.3.4.2.2.1). An element S of $M(3)$ may be written $S = [R, t]$ with $R \in O(3)$, the orthogonal group, and $t \in T(3)$, the translation group, and the product law

$$[R_1, t_1][R_2, t_2] = [R_1 R_2, t_1 + R_1(t_2)]$$

shows that $M(3) = T(3) \rtimes O(3)$ with $O(3)$ acting on $T(3)$ by rotating the translation vectors.

(*e*) *Associated actions in function spaces*

For every left action T_g of G in X, there is an associated left action $T_g^{\#}$ of G on the space $L(X)$ of complex-valued functions over X, defined by 'change of variable' (Section 1.3.2.3.9.5):

$$[T_g^{\#} f](x) = f((T_g)^{-1} x) = f(g^{-1} x).$$

Indeed for any g_1, g_2 in G,

$$[T_{g_1}^{\#}[T_{g_2}^{\#} f]](x) = [T_{g_2}^{\#} f]((T_{g_1})^{-1} x) = f[T_{g_2}^{-1} T_{g_1}^{-1} x]$$
$$= f((T_{g_1} T_{g_2})^{-1} x);$$

since $T_{g_1} T_{g_2} = T_{g_1 g_2}$, it follows that

$$T^{\#}_{g_1} T^{\#}_{g_2} = T^{\#}_{g_1 g_2}.$$

It is clear that the change of variable must involve the action of g^{-1} (not g) if $T^{\#}$ is to define a *left* action; using g instead would yield a *right* action.

The linear representation operators $T^{\#}_g$ on $L(X)$ provide the most natural instrument for stating and exploiting symmetry properties which a function may possess with respect to the action of G. Thus a function $f \in L(X)$ will be called *G-invariant* if $f(gx) = f(x)$ for all $g \in G$ and all $x \in X$. The value $f(x)$ then depends on x only through its orbit $G\,x$, and f is uniquely defined once it is specified on a fundamental domain $D = \{x_i\}_{i \in I}$; its integral over X is then a weighted sum of its values in D:

$$\sum_{x \in X} f(x) = \sum_{i \in I} [G : G_{x_i}] f(x_i).$$

The G-invariance of f may be written:

$$T^{\#}_g f = f \quad \text{for all } g \in G.$$

Thus f is invariant under each $T^{\#}_g$, which obviously implies that f is invariant under the linear operator in $L(X)$

$$A_G = \frac{1}{|G|} \sum_{g \in G} T^{\#}_g,$$

which averages an arbitrary function by the action of G. Conversely, if $A_G f = f$, then

$$T^{\#}_{g_0} f = T^{\#}_{g_0}(A_G f) = (T^{\#}_{g_0} A_G) f = A_G f = f \quad \text{for all } g_0 \in G,$$

so that the two statements of the G-invariance of f are equivalent. The identity

$$T^{\#}_{g_0} A_G = A_G \quad \text{for all } g_0 \in G$$

is easily proved by observing that the map $g \longmapsto g_0 g$ (g_0 being any element of G) is a one-to-one map from G into itself, so that

$$\sum_{g \in G} T^{\#}_g = \sum_{g \in G} T^{\#}_{g_0 g}$$

as these sums differ only by the order of the terms. The same identity implies that A_G is a *projector*:

$$(A_G)^2 = A_G,$$

and hence that its eigenvalues are either 0 or 1. In summary, we may say that the invariance of f under G is equivalent to f being an eigenfunction of the associated projector A_G for eigenvalue 1.

(f) *Orbit exchange*

One final result about group actions which will be used repeatedly later is concerned with the case when X has the structure of a Cartesian product:

$$X = X_1 \times X_2 \times \ldots \times X_n$$

and when G acts *diagonally* on X, *i.e.* acts on each X_j separately:

$$gx = g(x_1, x_2, \ldots, x_n) = (gx_1, gx_2, \ldots, gx_n).$$

Then complete sets (but not usually minimal sets) of representatives of the distinct orbits for the action of G in X may be obtained in the form

$$D_k = X_1 \times \ldots \times X_{k-1} \times \{x^{(k)}_{i_k}\}_{i_k \in I_k} \times X_{k+1} \times \ldots \times X_n$$

for each $k = 1, 2, \ldots, n$, *i.e.* by taking a fundamental domain in X_k and all the elements in X_j with $j \neq k$. The action of G on each D_k does indeed generate the whole of X: given an arbitrary element $y = (y_1, y_2, \ldots, y_n)$ of X, there is an index $i_k \in I_k$ such that $y_k \in G x^{(k)}_{i_k}$ and a coset of $G_{x^{(k)}_{i_k}}$ in G such that $y_k = \gamma x^{(k)}_{i_k}$ for any representative γ of that coset; then

$$y = \gamma(\gamma^{-1} y_1, \ldots, \gamma^{-1} y_{k-1}, x^{(k)}_{i_k}, \gamma^{-1} y_{k+1}, \ldots, \gamma^{-1} y_n)$$

which is of the form $y = \gamma d_k$ with $d_k \in D_k$.

The various D_k are related in a simple manner by 'transposition' or 'orbit exchange' (the latter name is due to J. W. Cooley). For instance, D_j may be obtained from $D_k (j \neq k)$ as follows: for each $y_j \in X_j$ there exists $g(y_j) \in G$ and $i_j(y_j) \in I_j$ such that $y_j = g(y_j) x^{(j)}_{i_j(y_j)}$; therefore

$$D_j = \bigcup_{y_j \in X_j} [g(y_j)]^{-1} D_k,$$

since the fundamental domain of X_k is thus expanded to the whole of X_k, while X_j is reduced to its fundamental domain. In other words: orbits are simultaneously collapsed in the jth factor and expanded in the kth.

When G operates without fixed points in each X_k (*i.e.* $G_{x_k} = \{e\}$ for all $x_k \in X_k$), then each D_k is a fundamental domain for the action of G in X. The existence of fixed points in some or all of the X_k complicates the situation in that for each k and each $x_k \in X_k$ such that $G_{x_k} \neq \{e\}$ the action of G/G_{x_k} on the other factors must be examined. Shenefelt (1988) has made a systematic study of orbit exchange for space group $P622$ and its subgroups.

Orbit exchange will be encountered, in a great diversity of forms, as the basic mechanism by which intermediate results may be rearranged between the successive stages of the computation of crystallographic Fourier transforms (Section 1.3.4.3).

1.3.4.2.2.3. Classification of crystallographic groups

Let Γ be a crystallographic group, Λ the normal subgroup of its lattice translations, and G the finite factor group Γ/Λ. Then G acts on Λ by conjugation [Section 1.3.4.2.2.2(d)] and this action, being a mapping of a lattice into itself, is representable by matrices with integer entries.

The classification of crystallographic groups proceeds from this observation in the following three steps:

Step 1: find all possible finite abstract groups G which can be represented by 3×3 integer matrices.

Step 2: for each such G find all its inequivalent representations by 3×3 integer matrices, equivalence being defined by a change of primitive lattice basis (*i.e.* conjugation by a 3×3 integer matrix with determinant ± 1).

Step 3: for each G and each equivalence class of integral representations of G, find all inequivalent extensions of the action of G from Λ to $T(3)$, equivalence being defined by an affine coordinate change [*i.e.* conjugation by an element of $A(3)$].

Step 1 leads to the following groups, listed in association with the crystal system to which they later give rise:

$\mathbb{Z}/2\mathbb{Z}$	monoclinic
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	orthorhombic
$\mathbb{Z}/3\mathbb{Z}, (\mathbb{Z}/3\mathbb{Z}) \ltimes\{\alpha\}$	trigonal
$\mathbb{Z}/4\mathbb{Z}, (\mathbb{Z}/4\mathbb{Z}) \ltimes\{\alpha\}$	tetragonal
$\mathbb{Z}/6\mathbb{Z}, (\mathbb{Z}/6\mathbb{Z}) \ltimes\{\alpha\}$	hexagonal
$(\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}) \ltimes\{S_3\}$	cubic

and the extension of these groups by a centre of inversion. In this list $\ltimes$ denotes a semi-direct product [Section 1.3.4.2.2.2(d)], α denotes the automorphism $g \longmapsto g^{-1}$, and S_3 (the group of permutations on three letters) operates by permuting the copies of $\mathbb{Z}/2\mathbb{Z}$ (using the subgroup A_3 of cyclic permutations gives the tetrahedral subsystem).

Step 2 leads to a list of 73 equivalence classes called *arithmetic classes* of representations $g \longmapsto \mathbf{R}_g$, where $\mathbf{R}_g$ is a 3×3 integer matrix, with $\mathbf{R}_{g_1 g_2} = \mathbf{R}_{g_1}\mathbf{R}_{g_2}$ and $\mathbf{R}_e = \mathbf{I}_3$. This enumeration is more familiar if equivalence is relaxed so as to allow conjugation by *rational* 3×3 matrices with determinant ± 1: this leads to the 32 crystal classes. The difference between an arithmetic class and its rational class resides in the choice of a lattice mode (P, $A/B/C$, I, F or R). Arithmetic classes always refer to a primitive lattice, but may use inequivalent integral representations for a given geometric symmetry element; while crystallographers prefer to change over to a nonprimitive lattice, if necessary, in order to preserve the same integral representation for a given geometric symmetry element. The matrices $\mathbf{P}$ and $\mathbf{Q} = \mathbf{P}^{-1}$ describing the changes of basis between primitive and centred lattices are listed in Table 5.1.3.1 and illustrated in Figs. 5.1.3.2 to 5.1.3.8, pp. 80–85, of Volume A of *International Tables* (Arnold, 2005).

Step 3 gives rise to a system of congruences for the systems of nonprimitive translations $\{\mathbf{t}_g\}_{g\in G}$ which may be associated to the matrices $\{\mathbf{R}_g\}_{g\in G}$ of a given arithmetic class, namely:

$$\mathbf{t}_{g_1 g_2} \equiv \mathbf{R}_{g_1}\mathbf{t}_{g_2} + \mathbf{t}_{g_1} \bmod \Lambda,$$

first derived by Frobenius (1911). If equivalence under the action of $A(3)$ is taken into account, 219 classes are found. If equivalence is defined with respect to the action of the subgroup $A^+(3)$ of $A(3)$ consisting only of transformations with determinant +1, then 230 classes called *space-group types* are obtained. In particular, associating to each of the 73 arithmetic classes a trivial set of nonprimitive translations ($\mathbf{t}_g = \mathbf{0}$ for all $g \in G$) yields the 73 symmorphic space groups. This third step may also be treated as an abstract problem concerning group extensions, using cohomological methods [Ascher & Janner (1965); see Janssen (1973) for a summary]; the connection with Frobenius's approach, as generalized by Zassenhaus (1948), is examined in Ascher & Janner (1968).

The finiteness of the number of space-group types in dimension 3 was shown by Bieberbach (1912) to be the case in arbitrary dimension. The reader interested in N-dimensional space-group theory for $N > 3$ may consult Brown (1969), Brown *et al.* (1978), Schwarzenberger (1980) and Engel (1986). The standard reference for integral representation theory is Curtis & Reiner (1962).

All three-dimensional space groups G have the property of being *solvable*, *i.e.* that there exists a chain of subgroups

$$G = G_r > G_{r-1} > \ldots > G_1 > G_0 = \{e\},$$

where each G_{i-1} is a normal subgroup of G_1 and the factor group G_i/G_{i-1} is a *cyclic* group of some order m_i ($1 \le i \le r$). This property may be established by inspection, or deduced from a famous theorem of Burnside [see Burnside (1911), pp. 322–323] according to which any group G such that $|G| = p^\alpha q^\beta$, with p and q distinct primes, is solvable; in the case at hand, $p = 2$ and $q = 3$. The whole classification of 3D space groups can be performed swiftly by a judicious use of the solvability property (L. Auslander, personal communication).

Solvability facilitates the indexing of elements of G in terms of generators and relations (Coxeter & Moser, 1972; Magnus *et al.*, 1976) for the purpose of calculation. By definition of solvability, elements $g_1, g_2, \ldots, g_r$ may be chosen in such a way that the cyclic factor group G_i/G_{i-1} is generated by the coset $g_i G_{i-1}$. The set $\{g_1, g_2, \ldots, g_r\}$ is then a system of generators for G such that the defining relations [see Brown *et al.* (1978), pp. 26–27] have the particularly simple form

$$\begin{aligned} g_1^{m_1} &= e, \\ g_i^{m_i} &= g_{i-1}^{a(i,i-1)} g_{i-2}^{a(i,i-2)} \ldots g_1^{a(i,1)} && \text{for } 2 \le i \le r, \\ g_i^{-1} g_j^{-1} g_i g_j &= g_{j-1}^{b(i,j,j-1)} g_{j-2}^{b(i,j,j-2)} \ldots g_1^{b(i,j,1)} && \text{for } 1 \le i < j \le r, \end{aligned}$$

with $0 \le a(i,h) < m_h$ and $0 \le b(i,j,h) < m_h$. Each element g of G may then be obtained uniquely as an 'ordered word':

$$g = g_r^{k_r} g_{r-1}^{k_{r-1}} \ldots g_1^{k_1},$$

with $0 \le k_i < m_i$ for all $i = 1, \ldots, r$, using the algorithm of Jürgensen (1970). Such generating sets and defining relations are tabulated in Brown *et al.* (1978, pp. 61–76). An alternative list is given in Janssen (1973, Table 4.3, pp. 121–123, and Appendix D, pp. 262–271).

1.3.4.2.2.4. Crystallographic group action in real space

The action of a crystallographic group Γ may be written in terms of standard coordinates in $\mathbb{R}^3/\mathbb{Z}^3$ as

$$(g, \mathbf{x}) \longmapsto S_g(\mathbf{x}) = \mathbf{R}_g\mathbf{x} + \mathbf{t}_g \bmod \Lambda, \quad g \in G,$$

with

$$S_{g_1 g_2} = S_{g_1} S_{g_2}.$$

An important characteristic of the representation $\theta : g \longmapsto S_g$ is its *reducibility*, *i.e.* whether or not it has invariant subspaces other than $\{\mathbf{0}\}$ and the whole of $\mathbb{R}^3/\mathbb{Z}^3$. For triclinic, monoclinic and orthorhombic space groups, θ is reducible to a direct sum of three one-dimensional representations:

$$\mathbf{R}_g = \begin{pmatrix} \mathbf{R}_g^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_g^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_g^{(3)} \end{pmatrix};$$

for trigonal, tetragonal and hexagonal groups, it is reducible to a direct sum of two representations, of dimension 2 and 1, respectively; while for tetrahedral and cubic groups, it is irreducible.

By Schur's lemma (see *e.g.* Ledermann, 1987), any matrix which commutes with all the matrices $\mathbf{R}_g$ for $g \in G$ must be a scalar multiple of the identity in each invariant subspace.

In the reducible cases, the reductions involve changes of basis which will be *rational*, not integral, for those arithmetic classes corresponding to nonprimitive lattices. Thus the simplification of having maximally reduced representation has as its counterpart the use of nonprimitive lattices.

The notions of orbit, isotropy subgroup and fundamental domain (or asymmetric unit) for the action of G on $\mathbb{R}^3/\mathbb{Z}^3$ are inherited directly from the general setting of Section 1.3.4.2.2.2. Points $\mathbf{x}$ for which $G_{\mathbf{x}} \neq \{e\}$ are called *special positions*, and the various types of isotropy subgroups which may be encountered in crystallographic groups have been labelled by means of Wyckoff symbols. The representation operators $S_g^{\#}$ in $L(\mathbb{R}^3/\mathbb{Z}^3)$ have the form:

$$[S_g^{\#} f](\mathbf{x}) = f\,[S_g^{-1}(\mathbf{x})] = f\,[\mathbf{R}_g^{-1}(\mathbf{x} - \mathbf{t}_g)].$$

The operators $R_g^{\#}$ associated to the purely rotational part of each transformation S_g will also be used. Note the relation: $S_g^{\#} = \tau_{\mathbf{t}_g} R_g^{\#}$.

Let a crystal structure be described by the list of the atoms in its unit cell, indexed by $k \in K$. Let the electron-density distribution about the centre of mass of atom k be described by ρ_k with respect to the standard coordinates $\mathbf{x}$. Then the motif ρ^0 may be written as a sum of translates:

$$\rho^0 = \sum_{k \in K} \tau_{\mathbf{x}_k} \rho_k$$

and the crystal electron density is $\rho = r^* \rho^0$.

Suppose that ρ is invariant under Γ. If $\mathbf{x}_{k_1}$ and $\mathbf{x}_{k_2}$ are in the same orbit, say $\mathbf{x}_{k_2} = S_g(\mathbf{x}_{k_1})$, then

$$\tau_{\mathbf{x}_{k_2}} \rho_{k_2} = S_g^{\#}(\tau_{\mathbf{x}_{k_1}} \rho_{k_1}).$$

Therefore if $\mathbf{x}_k$ is a special position and thus $G_{\mathbf{x}_k} \neq \{e\}$, then

$$S_g^{\#}(\tau_{\mathbf{x}_k} \rho_k) = \tau_{\mathbf{x}_k} \rho_k \quad \text{for all } g \in G_{\mathbf{x}_k}.$$

This identity implies that

$$\mathbf{R}_g \mathbf{x}_k + \mathbf{t}_g \equiv \mathbf{x}_k \bmod \Lambda$$

(the special position condition), and that

$$\rho_k = R_g^{\#} \rho_k,$$

i.e. that ρ_k must be invariant by the pure rotational part of $G_{\mathbf{x}_k}$. Trueblood (1956) investigated the consequences of this invariance on the thermal vibration tensor of an atom in a special position (see Section 1.3.4.2.2.6 below).

Let J be a subset of K such that $\{\mathbf{x}_j\}_{j \in J}$ contains exactly one atom from each orbit. An orbit decomposition yields an expression for ρ^0 in terms of symmetry-unique atoms:

$$\rho^0 = \sum_{j \in J} \left(\sum_{\gamma_j \in G/G_{\mathbf{x}_j}} S_{\gamma_j}^{\#}(\tau_{\mathbf{x}_j} \rho_j) \right)$$

or equivalently

$$\rho^0(\mathbf{x}) = \sum_{j \in J} \left\{ \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \rho_j[\mathbf{R}_{\gamma_j}^{-1}(\mathbf{x} - \mathbf{t}_{\gamma_j}) - \mathbf{x}_j] \right\}.$$

If the atoms are assumed to be Gaussian, write

$$\rho_j(\mathbf{X}) = \frac{Z_j}{|\det \pi \mathbf{U}_j|^{1/2}} \times \exp(-\tfrac{1}{2}\mathbf{X}^T \mathbf{U}_j^{-1} \mathbf{X}) \text{ in Cartesian Å coordinates,}$$

where Z_j is the total number of electrons, and where the matrix $\mathbf{U}_j$ combines the Gaussian spread of the electrons in atom j at rest with the covariance matrix of the random positional fluctuations of atom j caused by thermal agitation.

In crystallographic coordinates:

$$\rho_j(\mathbf{x}) = \frac{Z_j}{|\det \pi \mathbf{Q}_j|^{1/2}} \times \exp(-\tfrac{1}{2}\mathbf{x}^T \mathbf{Q}_j^{-1} \mathbf{x}) \text{ with } \mathbf{Q}_j = \mathbf{A}^{-1} \mathbf{U}_j (\mathbf{A}^{-1})^T.$$

If atom k is in a special position $\mathbf{x}_k$, then the matrix $\mathbf{Q}_k$ must satisfy the identity

$$\mathbf{R}_g \mathbf{Q}_k \mathbf{R}_g^{-1} = \mathbf{Q}_k$$

for all g in the isotropy subgroup of $\mathbf{x}_k$. This condition may also be written in Cartesian coordinates as

$$\mathbf{T}_g \mathbf{U}_k \mathbf{T}_g^{-1} = \mathbf{U}_k,$$

where

$$\mathbf{T}_g = \mathbf{A} \mathbf{R}_g \mathbf{A}^{-1}.$$

This is a condensed form of the symmetry properties derived by Trueblood (1956).

1.3.4.2.2.5. Crystallographic group action in reciprocal space

An elementary discussion of this topic may be found in Chapter 1.4 of this volume.

Having established that the symmetry of a crystal may be most conveniently stated and handled *via* the left representation $g \longmapsto S_g^{\#}$ of G given by its action on electron-density distributions, it is natural to transpose this action by the identity of Section 1.3.2.5.5:

$$\bar{\mathscr{F}}[S_g^{\#} T]_{\xi} = \bar{\mathscr{F}}[\tau_{\mathbf{t}_g}(R_g^{\#} T)]_{\xi} = \exp(2\pi i \boldsymbol{\xi} \cdot \mathbf{t}_g)[(\mathbf{R}_g^{-1})^{T\#} \bar{\mathscr{F}}[T]]_{\xi}$$

for any tempered distribution T, *i.e.*

$$\bar{\mathscr{F}}[S_g^{\#} T](\boldsymbol{\xi}) = \exp(2\pi i \boldsymbol{\xi} \cdot \mathbf{t}_g) \bar{\mathscr{F}}[T](\mathbf{R}_g^T \boldsymbol{\xi})$$

whenever the transforms are functions.

Putting $T = \rho$, a $\mathbb{Z}^3$-periodic distribution, this relation defines a left action S_g^* of G on $L(\mathbb{Z}^3)$ given by

$$(S_g^* F)(\mathbf{h}) = \exp(2\pi i \boldsymbol{\xi} \cdot \mathbf{t}_g) F(\mathbf{R}_g^T \mathbf{h})$$

which is conjugate to the action $S_g^{\#}$ in the sense that

$$\bar{\mathscr{F}}[S_g^{\#} \rho] = S_g^* \bar{\mathscr{F}}[\rho], \quad \textit{i.e. } S_g^* = \bar{\mathscr{F}} S_g^{\#} \bar{\mathscr{F}}.$$

The identity $S_g^{\#}\wp = \wp$ expressing the G-invariance of $\wp$ is then equivalent to the identity $S_g^* F = F$ between its structure factors, *i.e.* (Waser, 1955*a*)

$$F(\mathbf{h}) = \exp(2\pi i \mathbf{h}\cdot\mathbf{t}_g) F(\mathbf{R}_g^T \mathbf{h}).$$

If G is made to act on $\mathbb{Z}^3$ *via*

$$\theta^* : \quad (g, \mathbf{h}) \longmapsto (\mathbf{R}_g^{-1})^T \mathbf{h},$$

the usual notions of orbit, isotropy subgroup (denoted $G_{\mathbf{h}}$) and fundamental domain may be attached to this action. The above relation then shows that the spectrum $\{F(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^3}$ is entirely known if it is specified on a fundamental domain D^* containing one reciprocal-lattice point from each orbit of this action.

A reflection $\mathbf{h}$ is called *special* if $G_{\mathbf{h}} \neq \{e\}$. Then for any $g \in G_{\mathbf{h}}$ we have $\mathbf{R}_g^T \mathbf{h} = \mathbf{h}$, and hence

$$F(\mathbf{h}) = \exp(2\pi i \mathbf{h}\cdot\mathbf{t}_g) F(\mathbf{h}),$$

implying that $F(\mathbf{h}) = 0$ unless $\mathbf{h}\cdot\mathbf{t}_g \equiv 0 \bmod 1$. Special reflections $\mathbf{h}$ for which $\mathbf{h}\cdot\mathbf{t}_g \not\equiv 0 \bmod 1$ for some $g \in G_{\mathbf{h}}$ are thus *systematically absent*. This phenomenon is an instance of the duality between periodization and decimation of Section 1.3.2.7.2: if $\mathbf{t}_g \neq \mathbf{0}$, the projection of $\wp$ on the direction of $\mathbf{h}$ has period $(\mathbf{t}_g\cdot\mathbf{h})/(\mathbf{h}\cdot\mathbf{h}) < 1$, hence its transform (which is the portion of F supported by the central line through $\mathbf{h}$) will be decimated, giving rise to the above condition.

A reflection $\mathbf{h}$ is called *centric* if $G\mathbf{h} = G(-\mathbf{h})$, *i.e.* if the orbit of $\mathbf{h}$ contains $-\mathbf{h}$. Then $\mathbf{R}_\gamma^T \mathbf{h} = -\mathbf{h}$ for some coset γ in $G/G_{\mathbf{h}}$, so that the following relation must hold:

$$|F(\mathbf{h})| \exp(i\varphi_{\mathbf{h}}) = \exp(2\pi i \mathbf{h}\cdot\mathbf{t}_\gamma) |F(-\mathbf{h})| \exp(i\varphi_{-\mathbf{h}}).$$

In the absence of dispersion, Friedel's law gives rise to the *phase restriction*:

$$\varphi_{\mathbf{h}} \equiv \pi \mathbf{h}\cdot\mathbf{t}_\gamma \bmod \pi.$$

The value of the restricted phase is independent of the choice of coset representative γ. Indeed, if γ' is another choice, then $\gamma' = g\gamma$ with $g \in G_{\mathbf{h}}$ and by the Frobenius congruences $\mathbf{t}_{\gamma'} = \mathbf{R}_g \mathbf{t}_\gamma + \mathbf{t}_g$, so that

$$\mathbf{h}\cdot\mathbf{t}_{\gamma'} \equiv (\mathbf{R}_g^T \mathbf{h})\cdot\mathbf{t}_\gamma + \mathbf{h}\cdot\mathbf{t}_g \bmod 1.$$

Since $g \in G_{\mathbf{h}}$, $\mathbf{R}_g^T \mathbf{h} = \mathbf{h}$ and $\mathbf{h}\cdot\mathbf{t}_g \equiv 0 \bmod 1$ if $\mathbf{h}$ is not a systematic absence: thus

$$\pi \mathbf{h}\cdot\mathbf{t}_\gamma \equiv \pi \mathbf{h}\cdot\mathbf{t}_{\gamma'} \bmod \pi.$$

The treatment of centred lattices may be viewed as another instance of the duality between periodization and decimation (Section 1.3.2.7.2): the periodization of the electron density by the nonprimitive lattice translations has as its counterpart in reciprocal space the decimation of the transform by the 'reflection conditions' describing the allowed reflections, the decimation and periodization matrices being each other's contragredient.

The reader may consult the papers by Bienenstock & Ewald (1962) and Wells (1965) for earlier approaches to this material.

1.3.4.2.2.6. Structure-factor calculation

Structure factors may be calculated from a list of symmetry-unique atoms by Fourier transformation of the orbit decomposition formula for the motif $\wp^0$ given in Section 1.3.4.2.2.4:

$$\begin{aligned}
F(\mathbf{h}) &= \bar{\mathscr{F}}[\wp^0](\mathbf{h}) \\
&= \bar{\mathscr{F}}\left[\sum_{j\in J}\left(\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} S_{\gamma_j}^{\#}(\tau_{\mathbf{x}_j}\wp_j)\right)\right](\mathbf{h}) \\
&= \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \bar{\mathscr{F}}[\tau_{\mathbf{t}_{\gamma_j}} \mathbf{R}_{\gamma_j}^{\#} \tau_{\mathbf{x}_j}\wp_j](\mathbf{h}) \\
&= \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp(2\pi i \mathbf{h}\cdot\mathbf{t}_{\gamma_j}) \\
&\quad \times [(\mathbf{R}_{\gamma_j}^{-1})^{T\#}[\exp(2\pi i \boldsymbol{\xi}\cdot\mathbf{x}_j)\bar{\mathscr{F}}[\wp_j]_\xi]](\mathbf{h}) \\
&= \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp(2\pi i \mathbf{h}\cdot\mathbf{t}_{\gamma_j}) \\
&\quad \times \exp[2\pi i (\mathbf{R}_{\gamma_j}^T \mathbf{h})\cdot\mathbf{x}_j]\bar{\mathscr{F}}[\wp_j](\mathbf{R}_{\gamma_j}^T \mathbf{h});
\end{aligned}$$

i.e. finally:

$$F(\mathbf{h}) = \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp\{2\pi i \mathbf{h}\cdot[S_{\gamma_j}(\mathbf{x}_j)]\}\bar{\mathscr{F}}[\wp_j](\mathbf{R}_{\gamma_j}^T \mathbf{h}).$$

In the case of Gaussian atoms, the atomic transforms are

$$\bar{\mathscr{F}}[\wp_j](\mathbf{h}) = Z_j \exp[-\tfrac{1}{2}\mathbf{h}^T(4\pi^2\mathbf{Q}_j)\mathbf{h}]$$

or equivalently

$$\bar{\mathscr{F}}[\rho_j](\mathbf{H}) = Z_j \exp[-\tfrac{1}{2}\mathbf{H}^T(4\pi^2\mathbf{U}_j)\mathbf{H}].$$

Two common forms of equivalent temperature factors (incorporating both atomic form and thermal motion) are

(i) isotropic B:

$$\bar{\mathscr{F}}[\wp_j](\mathbf{h}) = Z_j \exp(-\tfrac{1}{4}B_j \mathbf{H}^T\mathbf{H}),$$

so that $\mathbf{U}_j = (B_j/8\pi^2)\mathbf{I}$, or $\mathbf{Q}_j = (B_j/8\pi^2)\mathbf{A}^{-1}(\mathbf{A}^{-1})^T$;

(ii) anisotropic β's:

$$\bar{\mathscr{F}}[\wp_j](\mathbf{h}) = Z_j \exp(-\mathbf{h}^T\boldsymbol{\beta}_j\mathbf{h}),$$

so that $\boldsymbol{\beta}_j = 2\pi^2\mathbf{Q}_j = 2\pi^2\mathbf{A}^{-1}\mathbf{U}_j(\mathbf{A}^{-1})^T$, or $\mathbf{U}_j = (1/2\pi^2)\mathbf{A}\beta_j\mathbf{A}^T$.

In the first case, $\bar{\mathscr{F}}[\wp_j](\mathbf{R}_{\gamma_j}^T\mathbf{h})$ does not depend on γ_j, and therefore:

$$\begin{aligned}
F(\mathbf{h}) &= \sum_{j\in J} Z_j \exp\{-\tfrac{1}{4}B_j \mathbf{h}^T[\mathbf{A}^{-1}(\mathbf{A}^{-1})^T]\mathbf{h}\} \\
&\quad \times \sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp\{2\pi i \mathbf{h}\cdot[S_{\gamma_j}(\mathbf{x}_j)]\}.
\end{aligned}$$

In the second case, however, no such simplification can occur:

$$\begin{aligned}
F(\mathbf{h}) &= \sum_{j\in J} Z_j \sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp[-\mathbf{h}^T(\mathbf{R}_{\gamma_j}\boldsymbol{\beta}_j\mathbf{R}_{\gamma_j}^T)\mathbf{h}] \\
&\quad \times \exp\{2\pi i \mathbf{h}\cdot[S_{\gamma_j}(\mathbf{x}_j)]\}.
\end{aligned}$$

These formulae, or special cases of them, were derived by Rollett & Davies (1955), Waser (1955*b*), and Trueblood (1956).

The computation of structure factors by applying the discrete Fourier transform to a set of electron-density values calculated on a grid will be examined in Section 1.3.4.4.5.

1.3.4.2.2.7. Electron-density calculations

A formula for the Fourier synthesis of electron-density maps from symmetry-unique structure factors is readily obtained by orbit decomposition:

$$\begin{aligned}\rho(\mathbf{x}) &= \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}) \\ &= \sum_{l\in L}\left[\sum_{\gamma_l\in G/G_{\mathbf{h}_l}} F(\mathbf{R}_{\gamma_l}^T\mathbf{h}_l)\exp[-2\pi i(\mathbf{R}_{\gamma_l}^T\mathbf{h}_l)\cdot\mathbf{x}]\right] \\ &= \sum_{l\in L} F(\mathbf{h}_l)\left[\sum_{\gamma_l\in G/G_{\mathbf{h}_l}}\exp\{-2\pi i\mathbf{h}_l\cdot[S_{\gamma_l}(\mathbf{x})]\}\right],\end{aligned}$$

where L is a subset of $\mathbb{Z}^3$ such that $\{\mathbf{h}_l\}_{l\in L}$ contains exactly one point of each orbit for the action $\theta^*:(g,\mathbf{h})\longmapsto(\mathbf{R}_g^{-1})^T\mathbf{h}$ of G on $\mathbb{Z}^3$. The physical electron density per cubic ångström is then

$$\rho(\mathbf{X}) = \frac{1}{V}\rho(\mathbf{A}\mathbf{x})$$

with V in Å^3.

In the absence of anomalous scatterers in the crystal and of a centre of inversion $-\mathbf{I}$ in Γ, the spectrum $\{F(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^3}$ has an extra symmetry, namely the Hermitian symmetry expressing Friedel's law (Section 1.3.4.2.1.4). The action of a centre of inversion may be added to that of Γ to obtain further simplification in the above formula: under this extra action, an orbit $G\mathbf{h}_l$ with $\mathbf{h}_l\neq\mathbf{0}$ is either mapped into itself or into the disjoint orbit $G(-\mathbf{h}_l)$; the terms corresponding to $+\mathbf{h}_l$ and $-\mathbf{h}_l$ may then be grouped *within* the common orbit in the first case, and *between* the two orbits in the second case.

Case 1: $G(-\mathbf{h}_l)=G\mathbf{h}_l$, $\mathbf{h}_l$ is *centric*. The cosets in $G/G_{\mathbf{h}_l}$ may be partitioned into two disjoint classes by picking one coset in each of the two-coset orbits of the action of $-\mathbf{I}$. Let $(G/G_{\mathbf{h}_l})^+$ denote one such class: then the *reduced orbit*

$$\{\mathbf{R}_{\gamma_l}^T\mathbf{h}_l|\gamma_l\in(G/G_{\mathbf{h}_l})^+\}$$

contains exactly once the Friedel-unique half of the full orbit $G\mathbf{h}_l$, and thus

$$|(G/G_{\mathbf{h}_l})^+| = \tfrac{1}{2}|G/G_{\mathbf{h}_l}|.$$

Grouping the summands for $+\mathbf{h}_l$ and $-\mathbf{h}_l$ yields a real-valued summand

$$2F(\mathbf{h}_l)\sum_{\gamma_l\in(G/G_{\mathbf{h}_l})^+}\cos[2\pi\mathbf{h}_l\cdot[S_{\gamma_l}(\mathbf{x})]-\varphi_{\mathbf{h}l}].$$

Case 2: $G(-\mathbf{h}_l)\neq G\mathbf{h}_l$, $\mathbf{h}_l$ is *acentric*. The two orbits are then disjoint, and the summands corresponding to $+\mathbf{h}_l$ and $-\mathbf{h}_l$ may be grouped together into a single real-valued summand

$$2F(\mathbf{h}_l)\sum_{\gamma_l\in G/G_{\mathbf{h}_l}}\cos[2\pi\mathbf{h}_l\cdot[S_{\gamma_l}(\mathbf{x})]-\varphi_{\mathbf{h}_l}].$$

In order to reindex the collection of all summands of ρ, put

$$L = L_c\cup L_a,$$

where L_c labels the Friedel-unique centric reflections in L and L_a the acentric ones, and let L_a^+ stand for a subset of L_a containing a unique element of each pair $\{+\mathbf{h}_l,-\mathbf{h}_l\}$ for $l\in L_a$. Then

$$\begin{aligned}\rho(\mathbf{x}) &= F(\mathbf{0}) \\ &\quad+\sum_{c\in L_c}\left[2F(\mathbf{h}_c)\sum_{\gamma_c\in(G/G_{\mathbf{h}_c})^+}\cos[2\pi\mathbf{h}_c\cdot[S_{\gamma_c}(\mathbf{x})]-\varphi_{\mathbf{h}_c}]\right] \\ &\quad+\sum_{a\in L_a^+}\left[2F(\mathbf{h}_a)\sum_{\gamma_a\in G/G_{\mathbf{h}_a}}\cos[2\pi\mathbf{h}_a\cdot[S_{\gamma_a}(\mathbf{x})]-\varphi_{\mathbf{h}_a}]\right].\end{aligned}$$

1.3.4.2.2.8. Parseval's theorem with crystallographic symmetry

The general statement of Parseval's theorem given in Section 1.3.4.2.1.5 may be rewritten in terms of symmetry-unique structure factors and electron densities by means of orbit decomposition.

In reciprocal space,

$$\sum_{\mathbf{h}\in\mathbb{Z}^3}\overline{F_1(\mathbf{h})}F_2(\mathbf{h}) = \sum_{l\in L}\sum_{\gamma_l\in G/G_{\mathbf{h}_l}}\overline{F_1(\mathbf{R}_{\gamma_l}^T\mathbf{h}_l)}F_2(\mathbf{R}_{\gamma_l}^T\mathbf{h}_l);$$

for each l, the summands corresponding to the various γ_l are equal, so that the left-hand side is equal to

$$\begin{aligned}&F_1(\mathbf{0})F_2(\mathbf{0}) \\ &\quad+\sum_{c\in L_c}2|(G/G_{\mathbf{h}_c})^+|\|F_1(\mathbf{h}_c)\|F_2(\mathbf{h}_c)|\cos[\varphi_1(\mathbf{h}_c)-\varphi_2(\mathbf{h}_c)] \\ &\quad+\sum_{a\in L_a^+}2|G/G_{\mathbf{h}_a}\|F_1(\mathbf{h}_a)\|F_2(\mathbf{h}_a)|\cos[\varphi_1(\mathbf{h}_a)-\varphi_2(\mathbf{h}_a)].\end{aligned}$$

In real space, the triple integral may be rewritten as

$$\int_{\mathbb{R}^3/\mathbb{Z}^3}\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x})\,\mathrm{d}^3\mathbf{x} = |G|\int_D\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$$

(where D is the asymmetric unit) if ρ_1 and ρ_2 are smooth densities, since the set of special positions has measure zero. If, however, the integral is approximated as a sum over a G-invariant grid defined by decimation matrix $\mathbf{N}$, special positions on this grid must be taken into account:

$$\begin{aligned}&\frac{1}{|\mathbf{N}|}\sum_{\mathbf{k}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3}\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x}) \\ &\quad=\frac{1}{|\mathbf{N}|}\sum_{\mathbf{x}\in D}[G:G_{\mathbf{x}}]\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x}) \\ &\quad=\frac{|G|}{|\mathbf{N}|}\sum_{\mathbf{x}\in D}\frac{1}{|G_{\mathbf{x}}|}\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x}),\end{aligned}$$

where the discrete asymmetric unit D contains exactly one point in each orbit of G in $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$.

1.3.4.2.2.9. Convolution theorems with crystallographic symmetry

The standard convolution theorems derived in the absence of symmetry are readily seen to follow from simple properties of functions $e^{\pm}(\mathbf{h},\mathbf{x}) = \exp(\pm 2\pi i \mathbf{h}\cdot\mathbf{x})$ (denoted simply e in formulae which are valid for both signs), namely:

$$\text{(i)} \qquad e(\mathbf{h},\mathbf{x}) \times e(\mathbf{k},\mathbf{x}) = e(\mathbf{h}+\mathbf{k},\mathbf{x}),$$
$$\text{(ii)} \qquad e(\mathbf{h},\mathbf{x}) \times e(\mathbf{h},\mathbf{y}) = e(\mathbf{h},\mathbf{x}+\mathbf{y}).$$

These relations imply that the families of functions

$$\{\mathbf{x} \longmapsto e(\mathbf{h},\mathbf{x})\}_{\mathbf{h}\in\mathbb{Z}^3} \qquad \text{in real space}$$

and

$$\{\mathbf{h} \longmapsto e(\mathbf{h},\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^3/\mathbb{Z}^3} \qquad \text{in reciprocal space}$$

both generate an *algebra* of functions, *i.e.* a vector space endowed with an internal multiplication, since (i) and (ii) show how to 'linearize products'.

Friedel's law (when applicable) on the one hand, and the Fourier relation between intensities and the Patterson function on the other hand, both follow from the property

$$\text{(iii)} \qquad \overline{e(\mathbf{h},\mathbf{x})} = e(-\mathbf{h},\mathbf{x}) = e(\mathbf{h},-\mathbf{x}).$$

When crystallographic symmetry is present, the convolution theorems remain valid in their original form if written out in terms of 'expanded' data, but acquire a different form when rewritten in terms of symmetry-unique data only. This rewriting is made possible by the extra relation (Section 1.3.4.2.2.5)

$$\text{(iv)} \qquad S^{\#}_{g^{-1}} e(\mathbf{h},\mathbf{x}) \equiv e[\mathbf{h}, S_g(\mathbf{x})] = e(\mathbf{h},\mathbf{t}_g) e(\mathbf{R}_g^T\mathbf{h},\mathbf{x})$$

or equivalently

$$\text{(iv}') \qquad S^{\#}_{g} e(\mathbf{h},\mathbf{x}) \equiv e[\mathbf{h}, S_g^{-1}(\mathbf{x})] = e[(-\mathbf{R}_g^{-1})^T\mathbf{h},\mathbf{t}_g] e[(\mathbf{R}_g^{-1})^T\mathbf{h},\mathbf{x}].$$

The kernels of symmetrized Fourier transforms are not the functions e but rather the symmetrized sums

$$\Xi^{\pm}(\mathbf{h},\mathbf{x}) = \sum_{g\in G} e^{\pm}[\mathbf{h}, S_g(\mathbf{x})] = \sum_{g\in G} e^{\pm}[\mathbf{h}, S_g^{-1}(\mathbf{x})]$$

for which the linearization formulae are readily obtained using (i), (ii) and (iv) as

$$\text{(i)}_G \qquad \Xi^{\pm}(\mathbf{h},\mathbf{x})\Xi^{\pm}(\mathbf{k},\mathbf{x}) = \sum_{g\in G} e^{\pm}(\mathbf{k},\mathbf{t}_g)\Xi^{\pm}(\mathbf{h}+\mathbf{R}_g^T\mathbf{k},\mathbf{x}),$$
$$\text{(ii)}_G \qquad \Xi^{\pm}(\mathbf{h},\mathbf{x})\Xi^{\pm}(\mathbf{h},\mathbf{y}) = \sum_{g\in G} \Xi^{\pm}[\mathbf{h},\mathbf{x}+S_g(\mathbf{y})],$$

where the choice of sign in $\pm$ must be the same throughout each formula.

Formulae $(\text{i})_G$ defining the 'structure-factor algebra' associated to G were derived by Bertaut (1955*c*, 1956*b*,*c*, 1959*a*,*b*) and Bertaut & Waser (1957) in another context.

The forward convolution theorem (in discrete form) then follows. Let

$$F_1(\mathbf{h}) = \sum_{\mathbf{y}\in D} \frac{1}{|G_{\mathbf{y}}|} \rho_1(\mathbf{y})\Xi^{+}(\mathbf{h},\mathbf{y}),$$
$$F_2(\mathbf{h}) = \sum_{\mathbf{z}\in D} \frac{1}{|G_{\mathbf{z}}|} \rho_2(\mathbf{z})\Xi^{+}(\mathbf{h},\mathbf{z}),$$

then

$$F_1(\mathbf{h})F_2(\mathbf{h}) = \sum_{\mathbf{x}\in D} \frac{1}{|G_{\mathbf{x}}|} \sigma(\mathbf{x})\Xi^{+}(\mathbf{h},\mathbf{x})$$

with

$$\sigma(\mathbf{x}) = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{z}\in D} \sum_{g\in G} \frac{|G_{\mathbf{x}}|}{|G_{\mathbf{x}-S_g(\mathbf{z})}| \times |G_{\mathbf{z}}|} \rho_1[\mathbf{x}-S_g(\mathbf{z})]\rho_2(\mathbf{z}).$$

The backward convolution theorem is derived similarly. Let

$$\rho_1(\mathbf{x}) = \sum_{\mathbf{k}\in D^*} \frac{1}{|G_{\mathbf{k}}|} F_1(\mathbf{k})\Xi^{-}(\mathbf{k},\mathbf{x}),$$
$$\rho_2(\mathbf{x}) = \sum_{\mathbf{l}\in D^*} \frac{1}{|G_{\mathbf{l}}|} F_2(\mathbf{l})\Xi^{-}(\mathbf{l},\mathbf{x}),$$

then

$$\rho_1(\mathbf{x})\rho_2(\mathbf{x}) = \sum_{\mathbf{h}\in D^*} \frac{1}{|G_{\mathbf{h}}|} F(\mathbf{h})\Xi^{-}(\mathbf{h},\mathbf{x})$$

with

$$F(\mathbf{h}) = \sum_{\mathbf{l}\in D^*} \sum_{g\in G} \frac{|G_{\mathbf{h}}|}{|G_{\mathbf{h}-\mathbf{R}_g^T(\mathbf{l})}| \times |G_{\mathbf{l}}|} e^{-}(\mathbf{l},\mathbf{t}_g) F_1(\mathbf{h}-\mathbf{R}_g^T\mathbf{l}) F_2(\mathbf{l}).$$

Both formulae are simply orbit decompositions of their symmetry-free counterparts.

1.3.4.2.2.10. Correlation and Patterson functions

Consider two model electron densities ρ_1 and ρ_2 with the same period lattice $\mathbb{Z}^3$ and the same space group G. Write their motifs in terms of atomic electron densities (Section 1.3.4.2.2.4) as

$$\rho_1^0 = \sum_{j_1\in J_1} \left(\sum_{\gamma_{j_1}\in G/G^{(1)}_{\mathbf{x}_{j_1}}} S^{\#}_{\gamma_{j_1}} (\tau_{\mathbf{x}^{(1)}_{j_1}} \rho^{(1)}_{j_1}) \right),$$
$$\rho_2^0 = \sum_{j_2\in J_2} \left(\sum_{\gamma_{j_2}\in G/G^{(2)}_{\mathbf{x}_{j_2}}} S^{\#}_{\gamma_{j_2}} (\tau_{\mathbf{x}^{(2)}_{j_2}} \rho^{(2)}_{j_2}) \right),$$

where J_1 and J_2 label the symmetry-unique atoms placed at positions $\{\mathbf{x}^{(1)}_{j_1}\}_{j_1\in J_1}$ and $\{\mathbf{x}^{(2)}_{j_2}\}_{j_2\in J_2}$, respectively.

To calculate the correlation between ρ_1 and ρ_2 we need the following preliminary formulae, which are easily established: if $S(\mathbf{x}) = \mathbf{R}\mathbf{x}+\mathbf{t}$ and f is an arbitrary function on $\mathbb{R}^3$, then

$$(R^{\#}f)\check{} = R^{\#}\check{f}, \quad (\tau_{\mathbf{x}}f)\check{} = \tau_{-\mathbf{x}}\check{f}, \quad R^{\#}(\tau_{\mathbf{x}}f) = \tau_{\mathbf{R}\mathbf{x}}f,$$

hence

$$S^{\#}(\tau_{\mathbf{x}} f) = \tau_{S(\mathbf{x})} R^{\#} f \quad \text{and} \quad [S^{\#}(\tau_{\mathbf{x}} f)]^{\check{}} = \tau_{-S(\mathbf{x})} R^{\#} \check{f};$$

and

$$S_1^{\#} f_1 * S_2^{\#} f_2 = S_1^{\#}[f_1 * (S_1^{-1} S_2)^{\#} f_2] = S_2^{\#}[(S_2^{-1} S_1)^{\#} f_1 * f_2].$$

The cross correlation $\check{\mathscr{P}}_1^0 * \mathscr{P}_2^0$ between motifs is therefore

$$\begin{aligned}\check{\mathscr{P}}_1^0 * \mathscr{P}_2^0 &= \sum_{j_1}\sum_{j_2}\sum_{\gamma_{j_1}}\sum_{\gamma_{j_2}} [S^{\#}_{\gamma_{j_1}}(\tau_{\mathbf{x}_{j_1}^{(1)}} \mathscr{P}_{j_1}^{(1)})]^{\check{}} * [S^{\#}_{\gamma_{j_2}}(\tau_{\mathbf{x}_{j_2}^{(2)}} \mathscr{P}_{j_2}^{(2)})] \\ &= \sum_{j_1}\sum_{j_2}\sum_{\gamma_{j_1}}\sum_{\gamma_{j_2}} \tau_{S_{\gamma_{j_2}}(\mathbf{x}_{j_2}^{(2)}) - S_{\gamma_{j_1}}(\mathbf{x}_{j_1}^{(1)})} [(R^{\#}_{\gamma_{j_1}} \check{\mathscr{P}}_{j_1}^{(1)}) * (R^{\#}_{\gamma_{j_2}} \mathscr{P}_{j_2}^{(2)})]\end{aligned}$$

which contains a peak of shape $(R^{\#}_{\gamma_{j_1}} \check{\mathscr{P}}_{j_1}^{(1)}) * (R^{\#}_{\gamma_{j_2}} \mathscr{P}_{j_2}^{(2)})$ at the interatomic vector $S_{\gamma_{j_2}}(\mathbf{x}_{j_2}^{(2)}) - S_{\gamma_{j_1}}(\mathbf{x}_{j_1}^{(1)})$ for each $j_1 \in J_1$, $j_2 \in J_2$, $\gamma_{j_1} \in G/G_{\mathbf{x}_{j_1}^{(1)}}$, $\gamma_{j_2} \in G/G_{\mathbf{x}_{j_2}^{(2)}}$.

The cross-correlation $r * \check{\mathscr{P}}^0 * \mathscr{P}^0$ between the original electron densities is then obtained by further periodizing by $\mathbb{Z}^3$.

Note that these expressions are valid for any choice of 'atomic' density functions $\mathscr{P}_{j_1}^{(1)}$ and $\mathscr{P}_{j_2}^{(2)}$, which may be taken as molecular fragments if desired (see Section 1.3.4.4.8).

If G contains elements g such that $\mathbf{R}_g$ has an eigenspace E_1 with eigenvalue 1 and an invariant complementary subspace E_2, while $\mathbf{t}_g$ has a nonzero component $\mathbf{t}_g^{(1)}$ in E_1, then the Patterson function $r * \check{\mathscr{P}}^0 * \mathscr{P}^0$ will contain *Harker peaks* (Harker, 1936) of the form

$$S_g(\mathbf{x}) - \mathbf{x} = \mathbf{t}_g^{(1)} \oplus (S_g^{(2)}(\mathbf{x}) - \mathbf{x})$$

[where $S_g^{(s)}$ represent the action of g in E_2] in the translate of E_1 by $\mathbf{t}_g^{(1)}$.

1.3.4.3. Crystallographic discrete Fourier transform algorithms

1.3.4.3.1. Historical introduction

In 1929, W. L. Bragg demonstrated the practical usefulness of the Fourier transform relation between electron density and structure factors by determining the structure of diopside from three principal projections calculated numerically by 2D Fourier summation (Bragg, 1929). It was immediately realized that the systematic use of this powerful method, and of its extension to three dimensions, would entail considerable amounts of numerical computation which had to be organized efficiently. As no other branch of applied science had yet needed this type of computation, crystallographers had to invent their own techniques.

The first step was taken by Beevers & Lipson (1934) who pointed out that a 2D summation could be factored into successive 1D summations. This is essentially the tensor product property of the Fourier transform (Sections 1.3.2.4.2.4, 1.3.3.3.1), although its aspect is rendered somewhat complicated by the use of sines and cosines instead of complex exponentials. Computation is economized to the extent that the cost of an $N \times N$ transform grows with N as $2N^3$ rather than N^4. Generalization to 3D is immediate, reducing computation size from N^6 to $3N^4$ for an $N \times N \times N$ transform. The complication introduced by using expressions in terms of sines and cosines is turned to advantage when symmetry is present, as certain families of terms are systematically absent or are simply related to each other; multiplicity corrections must, however, be introduced. The necessary information was tabulated for each space group by Lonsdale (1936), and was later incorporated into Volume I of *International Tables*.

The second step was taken by Beevers & Lipson (1936) and Lipson & Beevers (1936) in the form of the invention of the 'Beevers–Lipson strips', a practical device which was to assist a whole generation of crystallographers in the numerical computation of crystallographic Fourier sums. The strips comprise a set of 'cosine strips' tabulating the functions

$$A \cos\left(\frac{2\pi hm}{60}\right) \ (A = 1, 2, \ldots, 99;\ h = 1, 2, \ldots, 99)$$

and a set of 'sine strips' tabulating the functions

$$B \sin\left(\frac{2\pi hm}{60}\right) \ (B = 1, 2, \ldots, 99;\ h = 1, 2, \ldots, 99)$$

for the 16 arguments $m = 0, 1, \ldots, 15$. Function values are rounded to the nearest integer, and those for other arguments m may be obtained by using the symmetry properties of the sine and cosine functions. A Fourier summation of the form

$$Y(m) = \sum_{j=1}^{n} \left[A_j \cos\left(\frac{2\pi h_j m}{60}\right) + B_j \sin\left(\frac{2\pi h_j m}{60}\right) \right]$$

is then performed by selecting the n cosine strips labelled (A_j, h_j) and the n sine strips labelled (B_j, h_j), placing them in register, and adding the tabulated values columnwise. The number 60 was chosen as the l.c.m. of 12 (itself the l.c.m. of the orders of all possible nonprimitive translations) and of 10 (for decimal convenience). The limited accuracy imposed by the two-digit tabulation was later improved by Robertson's sorting board (Robertson, 1936*a*,*b*) or by the use of separate strips for each decimal digit of the amplitude (Booth, 1948*b*), which allowed three-digit tabulation while keeping the set of strips within manageable size. Cochran (1948*a*) found that, for most structures under study at the time, the numerical inaccuracies of the method were less than the level of error in the experimental data. The sampling rate was subsequently increased from 60 to 120 (Beevers, 1952) to cope with larger unit cells.

Further gains in speed and accuracy were sought through the construction of special-purpose mechanical, electro-mechanical, electronic or optical devices. Two striking examples are the mechanical computer RUFUS built by Robertson (1954, 1955, 1961) on the principle of previous strip methods (see also Robertson, 1932) and the electronic analogue computer X-RAC built by Pepinsky, capable of real-time calculation and display of 2D and 3D Fourier syntheses (Pepinsky, 1947; Pepinsky & Sayre, 1948; Pepinsky *et al.*, 1961; see also Suryan, 1957). The optical methods of Lipson & Taylor (1951, 1958) also deserve mention. Many other ingenious devices were invented, whose descriptions may be found in Booth (1948*b*), Niggli (1961) and Lipson & Cochran (1968).

Later, commercial punched-card machines were programmed to carry out Fourier summations or structure-factor calculations (Shaffer *et al.*, 1946*a*,*b*; Cox *et al.*, 1947, 1949; Cox & Jeffrey, 1949; Donohue & Schomaker, 1949; Grems & Kasper, 1949; Hodgson *et al.*, 1949; Greenhalgh & Jeffrey, 1950; Kitz & Marchington, 1953).

The modern era of digital electronic computation of Fourier series was initiated by the work of Bennett & Kendrew (1952), Mayer & Trueblood (1953), Ahmed & Cruickshank (1953*b*), Sparks *et al.* (1956) and Fowweather (1955). Their Fourier-synthesis programs used Beevers–Lipson factorization, the program by Sparks *et al.* being the first 3D Fourier program useable for all space groups (although these were treated as $P1$ or $P\bar{1}$ by data expansion). Ahmed & Barnes (1958) then proposed a

general programming technique to allow full use of symmetry elements (orthorhombic or lower) in the 3D Beevers–Lipson factorization process, including multiplicity corrections. Their method was later adopted by Shoemaker & Sly (1961), and by crystallographic program writers at large.

The discovery of the FFT algorithm by Cooley & Tukey in 1965, which instantly transformed electrical engineering and several other disciplines, paradoxically failed to have an immediate impact on crystallographic computing. A plausible explanation is that the calculation of large 3D Fourier maps was a relatively infrequent task which was not thought to constitute a bottleneck, as crystallographers had learned to settle most structural questions by means of cheaper 2D sections or projections. It is significant in this respect that the first use of the FFT in crystallography by Barrett & Zwick (1971) should have occurred as part of an iterative scheme for improving protein phases by density modification in real space, which required a much greater number of Fourier transformations than any previous method. Independently, Bondot (1971) had attracted attention to the merits of the FFT algorithm.

The FFT program used by Barrett & Zwick had been written for signal-processing applications. It was restricted to sampling rates of the form 2^n, and was not designed to take advantage of crystallographic symmetry at any stage of the calculation; Bantz & Zwick (1974) later improved this situation somewhat.

It was the work of Ten Eyck (1973) and Immirzi (1973, 1976) which led to the general adoption of the FFT in crystallographic computing. Immirzi treated all space groups as $P1$ by data expansion. Ten Eyck based his program on a versatile multi-radix FFT routine (Gentleman & Sande, 1966) coupled with a flexible indexing scheme for dealing efficiently with multidimensional transforms. He also addressed the problems of incorporating symmetry elements of order 2 into the factorization of 1D transforms, and of transposing intermediate results by other symmetry elements. He was thus able to show that in a large number of space groups (including the 74 space groups having orthorhombic or lower symmetry) it is possible to calculate only the unique results from the unique data within the logic of the FFT algorithm. Ten Eyck wrote and circulated a package of programs for computing Fourier maps and re-analysing them into structure factors in some simple space groups ($P1$, $P\bar{1}$, $P2$, $P2/m$, $P2_1$, $P222$, $P2_12_12_1$, $Pmmm$). This package was later augmented by a handful of new space-group-specific programs contributed by other crystallographers ($P2_12_12$, $I222$, $P3_121$, $P4_12_12$). The writing of such programs is an undertaking of substantial complexity, which has deterred all but the bravest: the usual practice is now to expand data for a high-symmetry space group to the largest subgroup for which a specific FFT program exists in the package, rather than attempt to write a new program. Attempts have been made to introduce more modern approaches to the calculation of crystallographic Fourier transforms (Auslander, Feig & Winograd, 1982; Auslander & Shenefelt, 1987; Auslander *et al.*, 1988) but have not gone beyond the stage of preliminary studies.

The task of fully exploiting the FFT algorithm in crystallographic computations is therefore still unfinished, and it is the purpose of this section to provide a systematic treatment such as that (say) of Ahmed & Barnes (1958) for the Beevers–Lipson algorithm.

Ten Eyck's approach, based on the reducibility of certain space groups, is extended by the derivation of a universal transposition formula for intermediate results. It is then shown that space groups which are not completely reducible may nevertheless be treated by three-dimensional Cooley–Tukey factorization in such a way that their symmetry may be fully exploited, whatever the shape of their asymmetric unit. Finally, new factorization methods with built-in symmetries are presented. The unifying concept throughout this presentation is that of 'group action' on indexing sets, and of 'orbit exchange' when this action has a composite structure; it affords new ways of rationalizing the use of symmetry, or of improving computational speed, or both.

1.3.4.3.2. Defining relations and symmetry considerations

A finite set of reflections $\{F_{\mathbf{h}_l}\}_{l\in L}$ can be periodized without aliasing by the translations of a suitable sublattice $\mathbf{N}^T\Lambda^*$ of the reciprocal lattice Λ^*; the converse operation in real space is the sampling of ρ at points $\mathbf{X}$ of a grid of the form $\mathbf{N}^{-1}\Lambda$ (Section 1.3.2.7.3). In standard coordinates, $\{F_{\mathbf{h}_l}\}_{l\in L}$ is periodized by $\mathbf{N}^T\mathbb{Z}^3$, and $\mathscr{P}$ is sampled at points $\mathbf{x}\in\mathbf{N}^{-1}\mathbb{Z}^3$.

In the absence of symmetry, the unique data are

– the $F_{\mathbf{h}}$ indexed by $\mathbf{h}\in\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3$ in reciprocal space;

– the $\mathscr{P}_{\mathbf{x}}$ indexed by $\mathbf{x}\in(\mathbf{N}^{-1}\mathbb{Z}^3)/\mathbb{Z}^3$; or equivalently the $\mathscr{P}_{\mathbf{m}}$ indexed by $\mathbf{m}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$, where $\mathbf{x}=\mathbf{N}^{-1}\mathbf{m}$.

They are connected by the ordinary DFT relations:

$$F_{\mathbf{h}}=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{x}\in(\mathbf{N}^{-1}\mathbb{Z}^3)/\mathbb{Z}^3}\mathscr{P}_{\mathbf{x}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x})$$

or

$$F_{\mathbf{h}}=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{m}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3}\mathscr{P}_{\mathbf{m}}\exp[2\pi i\mathbf{h}\cdot(\mathbf{N}^{-1}\mathbf{m})]$$

and

$$\mathscr{P}_{\mathbf{x}}=\sum_{\mathbf{h}=\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3}F_{\mathbf{h}}\exp(-2\pi i\mathbf{h}\cdot\mathbf{x})$$

or

$$\mathscr{P}_{\mathbf{m}}=\sum_{\mathbf{h}\in\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3}F_{\mathbf{h}}\exp[-2\pi i\mathbf{h}\cdot(\mathbf{N}^{-1}\mathbf{m})].$$

In the presence of symmetry, the unique data are

– $\{\mathscr{P}_{\mathbf{x}}\}_{\mathbf{x}\in D}$ or $\{\mathscr{P}_{\mathbf{m}}\}_{\mathbf{m}\in D}$ in real space (by abuse of notation, D will denote an asymmetric unit for $\mathbf{x}$ or for $\mathbf{m}$ indifferently);

– $\{F_{\mathbf{h}}\}_{\mathbf{h}\in D^*}$ in reciprocal space.

The previous summations may then be subjected to orbital decomposition, to yield the following 'crystallographic DFT' (CDFT) defining relations:

$$\begin{aligned}F_{\mathbf{h}}&=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{x}\in D}\mathscr{P}_{\mathbf{x}}\left[\sum_{\gamma\in G/G_{\mathbf{x}}}\exp\{2\pi i\mathbf{h}\cdot[S_\gamma(\mathbf{x})]\}\right]\\&=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{x}\in D}\mathscr{P}_{\mathbf{x}}\left[\frac{1}{|G_{\mathbf{x}}|}\sum_{g\in G}\exp\{2\pi i\mathbf{h}\cdot[S_g(\mathbf{x})]\}\right],\\\mathscr{P}_{\mathbf{x}}&=\sum_{\mathbf{h}\in D^*}F_{\mathbf{h}}\left[\sum_{\gamma\in G/G_{\mathbf{h}}}\exp\{-2\pi i\mathbf{h}\cdot[S_\gamma(\mathbf{x})]\}\right]\\&=\sum_{\mathbf{h}\in D^*}F_{\mathbf{h}}\left[\frac{1}{|G_{\mathbf{h}}|}\sum_{g\in G}\exp\{-2\pi i\mathbf{h}\cdot[S_g(\mathbf{x})]\}\right],\end{aligned}$$

with the obvious alternatives in terms of $\mathscr{P}_{\mathbf{m}}$, $\mathbf{m}=\mathbf{N}\mathbf{x}$. Our problem is to evaluate the CDFT for a given space group as efficiently as possible, in spite of the fact that the group action has spoilt the simple tensor-product structure of the ordinary three-dimensional DFT (Section 1.3.3.3.1).

Two procedures are available to carry out the 3D summations involved as a succession of smaller summations:

(1) *decomposition* into successive transforms of *fewer dimensions* but on the *same number of points* along these dimensions. This possibility depends on the reducibility of the space group, as defined in Section 1.3.4.2.2.4, and simply invokes the tensor product property of the DFT;

(2) *factorization* of the transform into transforms of the *same number of dimensions* as the original one, but on *fewer points* along each dimension. This possibility depends on the arithmetic factorability of the decimation matrix **N**, as described in Section 1.3.3.3.2.

Clearly, a symmetry expansion to the largest fully reducible subgroup of the space group will give maximal decomposability, but will require computing more than the unique results from more than the unique data. Economy will follow from factoring the transforms in the subspaces within which the space group acts irreducibly.

For irreducible subspaces of dimension 1, the group action is readily incorporated into the factorization of the transform, as first shown by Ten Eyck (1973).

For irreducible subspaces of dimension 2 or 3, the ease of incorporation of symmetry into the factorization depends on the type of factorization method used. The multidimensional Cooley–Tukey method (Section 1.3.3.3.1) is rather complicated; the multidimensional Good method (Section 1.3.3.3.2.2) is somewhat simpler; and the Rader/Winograd factorization admits a generalization, based on the arithmetic of certain rings of *algebraic* integers, which accommodates 2D crystallographic symmetries in a most powerful and pleasing fashion.

At each stage of the calculation, it is necessary to keep track of the definition of the asymmetric unit and of the symmetry properties of the numbers being manipulated. This requirement applies not only to the initial data and to the final results, where these are familiar; but also to all the intermediate quantities produced by partial transforms (on subsets of factors, or subsets of dimensions, or both), where they are less familiar. Here, the general formalism of transposition (or 'orbit exchange') described in Section 1.3.4.2.2.2 plays a central role.

1.3.4.3.3. Interaction between symmetry and decomposition

Suppose that the space-group action is reducible, *i.e.* that for each $g \in G$

$$\mathbf{R}_g = \begin{pmatrix} \mathbf{R}'_g & \mathbf{0} \\ \mathbf{0} & \mathbf{R}''_g \end{pmatrix}, \qquad \mathbf{t}_g = \begin{pmatrix} \mathbf{t}'_g \\ \mathbf{t}''_g \end{pmatrix};$$

by Schur's lemma, the decimation matrix must then be of the form

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}' & \mathbf{0} \\ \mathbf{0} & \mathbf{N}'' \end{pmatrix}$$

if it is to commute with all the $\mathbf{R}_g$.

Putting

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}' \\ \mathbf{x}'' \end{pmatrix} \quad \text{and} \quad \mathbf{h} = \begin{pmatrix} \mathbf{h}' \\ \mathbf{h}'' \end{pmatrix},$$

we may define

$$S'_g(\mathbf{x}') = \mathbf{R}'_g\mathbf{x}' + \mathbf{t}'_g,$$
$$S''_g(\mathbf{x}'') = \mathbf{R}''_g\mathbf{x}'' + \mathbf{t}''_g,$$

and write $S_g = S'_g \oplus S''_g$ (direct sum) as a shorthand for

$$S_g(\mathbf{x}) = \begin{pmatrix} S'_g(\mathbf{x}') \\ S''_g(\mathbf{x}'') \end{pmatrix}.$$

We may also define the representation operators $S'^{\#}_g$ and $S''^{\#}_g$ acting on functions of $\mathbf{x}'$ and $\mathbf{x}''$, respectively (as in Section 1.3.4.2.2.4), and the operators S'^{*}_g and S''^{*}_g acting on functions of $\mathbf{h}'$ and $\mathbf{h}''$, respectively (as in Section 1.3.4.2.2.5). Then we may write

$$S^{\#}_g = (S'_g)^{\#} \oplus (S''_g)^{\#}$$

and

$$S^{*}_g = (S'_g)^{*} \oplus (S''_g)^{*}$$

in the sense that g acts on $f(\mathbf{x}) \equiv f(\mathbf{x}', \mathbf{x}'')$ by

$$(S^{\#}_g f)(\mathbf{x}', \mathbf{x}'') = f[(S'_g)^{-1}(\mathbf{x}'), (S''_g)^{-1}(\mathbf{x}'')]$$

and on $\Phi(\mathbf{h}) \equiv \Phi(\mathbf{h}', \mathbf{h}'')$ by

$$(S^{*}_g \Phi)(\mathbf{h}', \mathbf{h}'') = \exp(2\pi i \mathbf{h}' \cdot \mathbf{t}'_g) \exp(2\pi i \mathbf{h}'' \cdot \mathbf{t}''_g) \times \Phi[\mathbf{R}'^{T}_g \mathbf{h}', \mathbf{R}''^{T}_g \mathbf{h}''].$$

Thus equipped we may now derive concisely a general identity describing the symmetry properties of intermediate quantities of the form

$$T(\mathbf{x}', \mathbf{h}'') = \sum_{\mathbf{h}'} F(\mathbf{h}', \mathbf{h}'') \exp(-2\pi i \mathbf{h}' \cdot \mathbf{x}') = \frac{1}{|\det \mathbf{N}'|} \sum_{\mathbf{x}''} \rho(\mathbf{x}', \mathbf{x}'') \exp(+2\pi i \mathbf{h}'' \cdot \mathbf{x}''),$$

which arise through partial transformation of F on $\mathbf{h}'$ or of ρ on $\mathbf{x}''$. The action of $g \in G$ on these quantities will be

(i) through $(S'_g)^{\#}$ on the function $\mathbf{x}' \longmapsto T(\mathbf{x}', \mathbf{h}'')$,

(ii) through $(S''_g)^{*}$ on the function $\mathbf{h}'' \longmapsto T(\mathbf{x}', \mathbf{h}'')$,

and hence the symmetry properties of T are expressed by the identity

$$T = [(S'_g)^{\#} \oplus (S''_g)^{*}]T.$$

Applying this relation not to T but to $[(S'_{g^{-1}})^{\#} \oplus (S''_e)^{*}]T$ gives

$$[(S'_{g^{-1}})^{\#} \oplus (S''_e)^{*}]T = [(S'_e)^{\#} \oplus (S''_g)^{*}]T,$$

i.e.

$$\boxed{T(S'_g(\mathbf{x}'), \mathbf{h}'') = \exp(2\pi i \mathbf{h}'' \cdot \mathbf{t}''_g) T(\mathbf{x}', \mathbf{R}''^{T}_g \mathbf{h}'').}$$

If the unique $F(\mathbf{h}) \equiv F(\mathbf{h}', \mathbf{h}'')$ were initially indexed by

$$(\text{all } \mathbf{h}') \times (\text{unique } \mathbf{h}'')$$

(see Section 1.3.4.2.2.2), this formula allows the reindexing of the intermediate results $T(\mathbf{x}', \mathbf{h}'')$ from the initial form

$$(\text{all } \mathbf{x}') \times (\text{unique } \mathbf{h}'')$$

to the final form

$$(\text{unique } \mathbf{x}') \times (\text{all } \mathbf{h}''),$$

on which the second transform (on $\mathbf{h}''$) may now be performed, giving the final results $\mathscr{P}(\mathbf{x}', \mathbf{x}'')$ indexed by

$$(\text{unique } \mathbf{x}') \times (\text{all } \mathbf{x}''),$$

which is an asymmetric unit. An analogous interpretation holds if one is going from $\mathscr{P}$ to F.

The above formula solves the general problem of transposing from one invariant subspace to another, and is the main device for decomposing the CDFT. Particular instances of this formula were derived and used by Ten Eyck (1973); it is useful for orthorhombic groups, and for dihedral groups containing screw axes n_m with g.c.d. $(m, n) = 1$. For comparison with later uses of orbit exchange, it should be noted that the type of intermediate results just dealt with is obtained after transforming on *all* factors in *one* summand.

A central piece of information for driving such a decomposition is the definition of the full asymmetric unit in terms of the asymmetric units in the invariant subspaces. As indicated at the end of Section 1.3.4.2.2.2, this is straightforward when G acts without fixed points, but becomes more involved if fixed points do exist. To this day, no systematic 'calculus of asymmetric units' exists which can automatically generate a complete description of the asymmetric unit of an arbitrary space group in a form suitable for directing the orbit exchange process, although Shenefelt (1988) has outlined a procedure for dealing with space group $P622$ and its subgroups. The asymmetric unit definitions given in Volume A of *International Tables* are incomplete in this respect, in that they do not specify the possible residual symmetries which may exist on the boundaries of the domains.

1.3.4.3.4. Interaction between symmetry and factorization

Methods for factoring the DFT in the absence of symmetry were examined in Sections 1.3.3.2 and 1.3.3.3. They are based on the observation that the finite sets which index both data and results are endowed with certain algebraic structures (*e.g.* are Abelian groups, or rings), and that subsets of indices may be found which are not merely subsets but *substructures* (*e.g.* subgroups or subrings). Summation over these substructures leads to partial transforms, and the way in which substructures fit into the global structure indicates how to reassemble the partial results into the final results. As a rule, the richer the algebraic structure which is identified in the indexing set, the more powerful the factoring method.

The ability of a given factoring method to accommodate crystallographic symmetry will thus be determined by the extent to which the crystallographic group action respects (or fails to respect) the partitioning of the index set into the substructures pertaining to that method. This remark justifies trying to gain an overall view of the algebraic structures involved, and of the possibilities of a crystallographic group acting 'naturally' on them.

The index sets $\{\mathbf{m}|\mathbf{m} \in \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3\}$ and $\{\mathbf{h}|\mathbf{h} \in \mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3\}$ are finite Abelian groups under component-wise addition. If an iterated addition is viewed as an action of an integer scalar $n \in \mathbb{Z}$ *via*

$$\begin{aligned} n\mathbf{h} &= \mathbf{h} + \mathbf{h} + \ldots + \mathbf{h} && (n \text{ times}) && \text{for } n > 0, \\ &= \mathbf{0} && && \text{for } n = 0, \\ &= -(\mathbf{h} + \mathbf{h} + \ldots + \mathbf{h}) && (|n| \text{ times}) && \text{for } n < 0, \end{aligned}$$

then an Abelian group becomes a *module* over the ring $\mathbb{Z}$ (or, for short, a $\mathbb{Z}$-module), a module being analogous to a vector space but with scalars drawn from a ring rather than a field. The left actions of a crystallographic group G by

$$g: \quad \mathbf{m} \longmapsto \mathbf{R}_g\mathbf{m} + \mathbf{N}\mathbf{t}_g \mod \mathbf{N}\mathbb{Z}^3$$

and by

$$g: \quad \mathbf{h} \longmapsto (\mathbf{R}_g^{-1})^T\mathbf{h} \mod \mathbf{N}^T\mathbb{Z}^3$$

can be combined with this $\mathbb{Z}$ action as follows:

$$\begin{aligned} \textstyle\sum_{g\in G} n_g g: &\quad \mathbf{m} \longmapsto \textstyle\sum_{g\in G} n_g(\mathbf{R}_g\mathbf{m} + \mathbf{N}\mathbf{t}_g) && \mod \mathbf{N}\mathbb{Z}^3, \\ \textstyle\sum_{g\in G} n_g g: &\quad \mathbf{h} \longmapsto \textstyle\sum_{g\in G} n_g[(\mathbf{R}_g^{-1})^T\mathbf{h}] && \mod \mathbf{N}^T\mathbb{Z}^3. \end{aligned}$$

This provides a left action, on the indexing sets, of the set

$$\mathbb{Z}G = \left\{ \sum_{g\in G} n_g g \middle| n_g \in \mathbb{Z} \text{ for each } g \in G \right\}$$

of symbolic linear combinations of elements of G with integral coefficients. If addition and multiplication are defined in $\mathbb{Z}G$ by

$$\left(\sum_{g_1\in G} a_{g_1} g_1\right) + \left(\sum_{g_2\in G} b_{g_2} g_2\right) = \sum_{g\in G} (a_g + b_g) g$$

and

$$\left(\sum_{g_1\in G} a_{g_1} g_1\right) \times \left(\sum_{g_2\in G} b_{g_2} g_2\right) = \sum_{g\in G} c_g g,$$

with

$$c_g = \sum_{g'\in G} a_{g'} b_{(g')^{-1}g},$$

then $\mathbb{Z}G$ is a *ring*, and the action defined above makes the indexing sets into $\mathbb{Z}G$*-modules*. The ring $\mathbb{Z}G$ is called the *integral group ring* of G (Curtis & Reiner, 1962, p. 44).

From the algebraic standpoint, therefore, the interaction between symmetry and factorization can be expected to be favourable whenever the indexing sets of partial transforms are $\mathbb{Z}G$-submodules of the main $\mathbb{Z}G$-modules.

1.3.4.3.4.1. Multidimensional Cooley–Tukey factorization

Suppose, as in Section 1.3.3.3.2.1, that the decimation matrix $\mathbf{N}$ may be factored as $\mathbf{N}_1\mathbf{N}_2$. Then any grid point index $\mathbf{m} \in \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$ in real space may be written

$$\mathbf{m} = \mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2$$

with $\mathbf{m}_1 \in \mathbb{Z}^3/\mathbf{N}_1\mathbb{Z}^3$ and $\mathbf{m}_2 \in \mathbb{Z}^3/\mathbf{N}_2\mathbb{Z}^3$ determined by

$$\begin{aligned} \mathbf{m}_1 &= \mathbf{m} && \mod \mathbf{N}_1\mathbb{Z}^3, \\ \mathbf{m}_2 &= \mathbf{N}_1^{-1}(\mathbf{m} - \mathbf{m}_1) && \mod \mathbf{N}_2\mathbb{Z}^3. \end{aligned}$$

These relations establish a one-to-one correspondence $\mathbf{m} \leftrightarrow (\mathbf{m}_1, \mathbf{m}_2)$ between $I = \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$ and the Cartesian product $I_1 \times I_2$ of $I_1 = \mathbb{Z}^3/\mathbf{N}_1\mathbb{Z}^3$ and $I_2 = \mathbb{Z}^3/\mathbf{N}_2\mathbb{Z}^3$, and hence $I \cong I_1 \times I_2$ as a set. However $I \ncong I_1 \times I_2$ as an Abelian group, since in general $\mathbf{m} + \mathbf{m}' \not\leftrightarrow (\mathbf{m}_1 + \mathbf{m}_1', \mathbf{m}_2 + \mathbf{m}_2')$ because there can be a 'carry' from the addition of the first components into the second components; therefore, $I \ncong I_1 \times I_2$ as a $\mathbb{Z}G$-module, which shows that the incorporation of symmetry into the Cooley–Tukey algorithm is not a trivial matter.

Let $g \in G$ act on I through

$$g: \quad \mathbf{m} \longmapsto S_g(\mathbf{m}) = \mathbf{R}_g\mathbf{m} + \mathbf{N}\mathbf{t}_g \bmod \mathbf{N}\mathbb{Z}^3$$

and suppose that $\mathbf{N}$ 'integerizes' all the nonprimitive translations $\mathbf{t}_g$ so that we may write

$$\mathbf{N}\mathbf{t}_g = \mathbf{t}_g^{(1)} + \mathbf{N}_1\mathbf{t}_g^{(2)},$$

with $\mathbf{t}_g^{(1)} \in I_1$ and $\mathbf{t}_g^{(2)} \in I_2$ determined as above. Suppose further that $\mathbf{N}$, $\mathbf{N}_1$ and $\mathbf{N}_2$ commute with $\mathbf{R}_g$ for all $g \in G$, *i.e.* (by Schur's lemma, Section 1.3.4.2.2.4) that these matrices are integer multiples of the identity in each G-invariant subspace. The action of g on $\mathbf{m} = \mathbf{N}\mathbf{x} \bmod \mathbf{N}\mathbb{Z}^3$ leads to

$$\begin{aligned} S_g(\mathbf{m}) &= \mathbf{N}[\mathbf{R}_g(\mathbf{N}^{-1}\mathbf{m}) + \mathbf{N}\mathbf{t}_g] && \bmod \mathbf{N}\mathbb{Z}^3 \\ &= \mathbf{N}\mathbf{R}_g\mathbf{N}^{-1}(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) + \mathbf{t}_g^{(1)} + \mathbf{N}_1\mathbf{t}_g^{(2)} && \bmod \mathbf{N}\mathbb{Z}^3 \\ &= \mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)} + \mathbf{N}_1(\mathbf{R}_g\mathbf{m}_2 + \mathbf{t}_g^{(2)}) && \bmod \mathbf{N}\mathbb{Z}^3, \end{aligned}$$

which we may decompose as

$$S_g(\mathbf{m}) = [S_g(\mathbf{m})]_1 + \mathbf{N}_1[S_g(\mathbf{m})]_2$$

with

$$[S_g(\mathbf{m})]_1 \equiv S_g(\mathbf{m}) \qquad \bmod \mathbf{N}_1\mathbb{Z}^3$$

and

$$[S_g(\mathbf{m})]_2 \equiv \mathbf{N}_1^{-1}\{S_g(\mathbf{m}) - [S_g(\mathbf{m})]_1\} \qquad \bmod \mathbf{N}_2\mathbb{Z}^3.$$

Introducing the notation

$$\begin{aligned} S_g^{(1)}(\mathbf{m}_1) &= \mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)} \bmod \mathbf{N}_1\mathbb{Z}^3, \\ S_g^{(2)}(\mathbf{m}_2) &= \mathbf{R}_g\mathbf{m}_2 + \mathbf{t}_g^{(2)} \bmod \mathbf{N}_2\mathbb{Z}^3, \end{aligned}$$

the two components of $S_g(\mathbf{m})$ may be written

$$\begin{aligned} [S_g(\mathbf{m})]_1 &= S_g^{(1)}(\mathbf{m}_1), \\ [S_g(\mathbf{m})]_2 &= S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^3, \end{aligned}$$

with

$$\boldsymbol{\mu}_2(g, \mathbf{m}_1) = \mathbf{N}_1^{-1}\{(\mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)}) - [S_g(\mathbf{m}_1)]_1\} \bmod \mathbf{N}_2\mathbb{Z}^3.$$

The term $\boldsymbol{\mu}_2$ is the geometric equivalent of a *carry* or *borrow*: it arises because $\mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)}$, calculated as a vector in $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$, may be outside the unit cell $\mathbf{N}_1[0, 1]^3$, and may need to be brought back into it by a 'large' translation with a nonzero component in the $\mathbf{m}_2$ space; equivalently, the action of g may need to be applied around different permissible origins for different values of $\mathbf{m}_1$, so as to map the unit cell into itself without any recourse to lattice translations. [Readers familiar with the cohomology of groups (see *e.g.* Hall, 1959; MacLane, 1963) will recognize $\boldsymbol{\mu}_2$ as the cocycle of the extension of $\mathbb{Z}$ G-modules described by the exact sequence $0 \to I_2 \to I \to I_1 \to 0$.]

Thus G acts on I in a rather complicated fashion: although $g \longmapsto S_g^{(1)}$ does define a left action in I_1 alone, no action can be defined in I_2 alone because $\boldsymbol{\mu}_2$ depends on $\mathbf{m}_1$. However, because S_g, $S_g^{(1)}$ and $S_g^{(2)}$ *are* left actions, it follows that $\boldsymbol{\mu}_2$ satisfies the identity

$$\boldsymbol{\mu}_2(gg', \mathbf{m}_1) = S_g^{(2)}[\boldsymbol{\mu}_2(g', \mathbf{m}_1)] + \boldsymbol{\mu}_2[g, S_g^{(1)}(\mathbf{m}_1)] \qquad \bmod \mathbf{N}_2\mathbb{Z}^3$$

for all g, g' in G and all $\mathbf{m}_1$ in I_1. In particular, $\boldsymbol{\mu}_2(\mathrm{e}, \mathbf{m}_1) = \mathbf{0}$ for all $\mathbf{m}_1$, and

$$\boldsymbol{\mu}_2(g^{-1}, \mathbf{m}_1) = -S_{g^{-1}}^{(2)}\{\boldsymbol{\mu}_2[g, S_{g^{-1}}^{(1)}(\mathbf{m}_1)]\} \bmod \mathbf{N}_2\mathbb{Z}^3.$$

This action will now be used to achieve optimal use of symmetry in the multidimensional Cooley–Tukey algorithm of Section 1.3.3.3.2.1. Let us form an array Y according to

$$Y(\mathbf{m}_1, \mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$$

for *all* $\mathbf{m}_2 \in I_2$ but only for the *unique* $\mathbf{m}_1$ under the action $S_g^{(1)}$ of G in I_1. Except in special cases which will be examined later, these vectors contain essentially an asymmetric unit of electron-density data, up to some redundancies on boundaries. We may then compute the partial transform on $\mathbf{m}_2$:

$$Y^*(\mathbf{m}_1, \mathbf{h}_2) = \frac{1}{|\det \mathbf{N}_2|} \sum_{\mathbf{m}_2 \in I_2} Y(\mathbf{m}_1, \mathbf{m}_2) e[\mathbf{h}_2 \cdot (\mathbf{N}_2^{-1}\mathbf{m}_2)].$$

Using the symmetry of ρ in the form $\rho = S_g^{\#}\rho$ yields by the procedure of Section 1.3.3.3.2 the transposition formula

$$\begin{aligned} Y^*(S_g^{(1)}(\mathbf{m}_1), \mathbf{h}_2) = e\{\mathbf{h}_2 \cdot [\mathbf{N}_2^{-1}(\mathbf{t}_g^{(2)} + \boldsymbol{\mu}_2(g, \mathbf{m}_1))]\} \\ \times Y^*(\mathbf{m}_1, [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2). \end{aligned}$$

By means of this identity we can transpose intermediate results Y^* initially indexed by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{h}_2),$$

so as to have them indexed by

$$(\text{all } \mathbf{m}_1) \times (\text{unique } \mathbf{h}_2).$$

We may then apply twiddle factors to get

$$Z(\mathbf{m}_1, \mathbf{h}_2) = e[\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] Y^*(\mathbf{m}_1, \mathbf{h}_2)$$

and carry out the second transform

$$Z^*(\mathbf{h}_1, \mathbf{h}_2) = \frac{1}{|\det \mathbf{N}_1|} \sum_{\mathbf{m}_1 \in I_1} Z(\mathbf{m}_1, \mathbf{h}_2) e[\mathbf{h}_1 \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)].$$

The final results are indexed by

$$(\text{all } \mathbf{h}_1) \times (\text{unique } \mathbf{h}_2),$$

which yield essentially an asymmetric unit of structure factors after unscrambling by:

$$F(\mathbf{h}_2 + \mathbf{N}_2^T\mathbf{h}_1) = Z^*(\mathbf{h}_1, \mathbf{h}_2).$$

The transposition formula above applies to intermediate results when going backwards from F to $\not\rho$, provided these results are considered *after* the twiddle-factor stage. A transposition formula applicable *before* that stage can be obtained by characterizing the action of G on $\mathbf{h}$ (including the effects of periodization by $\mathbf{N}^T\mathbb{Z}^3$) in a manner similar to that used for $\mathbf{m}$.

Let

$$\mathbf{h} = \mathbf{h}_2 + \mathbf{N}_2^T\mathbf{h}_1,$$

with

$$\begin{aligned} \mathbf{h}_2 &= \mathbf{h} && \bmod \mathbf{N}_2^T\mathbb{Z}^3, \\ \mathbf{h}_1 &= (\mathbf{N}_2^{-1})^T(\mathbf{h} - \mathbf{h}_2) && \bmod \mathbf{N}_1^T\mathbb{Z}^3. \end{aligned}$$

We may then write

$$\mathbf{R}_g^T\mathbf{h} = [\mathbf{R}_g^T\mathbf{h}]_2 + \mathbf{N}_2^T[\mathbf{R}_g^T\mathbf{h}]_1,$$

with

$$\begin{aligned} [\mathbf{R}_g^T\mathbf{h}]_2 &= [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2 && \bmod \mathbf{N}_2^T\mathbb{Z}^3, \\ [\mathbf{R}_g^T\mathbf{h}]_1 &= [\mathbf{R}_g^{(1)}]^T\mathbf{h}_1 + \boldsymbol{\eta}_1(g, \mathbf{h}_2) && \bmod \mathbf{N}_1^T\mathbb{Z}^3. \end{aligned}$$

Here $[\mathbf{R}_g^{(2)}]^T$, $[\mathbf{R}_g^{(1)}]^T$ and $\boldsymbol{\eta}_1$ are defined by

$$\begin{aligned} [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2 &= \mathbf{R}_g^T\mathbf{h} && \bmod \mathbf{N}_2^T\mathbb{Z}^3, \\ [\mathbf{R}_g^{(1)}]^T\mathbf{h}_1 &= \mathbf{R}_g^T\mathbf{h} && \bmod \mathbf{N}_1^T\mathbb{Z}^3 \end{aligned}$$

and

$$\boldsymbol{\eta}_1(g, \mathbf{h}_2) = (\mathbf{N}_2^{-1})^T(\mathbf{R}_g^T\mathbf{h}_2 - [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2) \bmod \mathbf{N}_1^T\mathbb{Z}^3.$$

Let us then form an array Z^* according to

$$Z^*(\mathbf{h}_1', \mathbf{h}_2') = F(\mathbf{h}_2' + \mathbf{N}_2^T\mathbf{h}_1')$$

for *all* $\mathbf{h}_1'$ but only for the *unique* $\mathbf{h}_2'$ under the action of G in $\mathbb{Z}^3/\mathbf{N}_2^T\mathbb{Z}^3$, and transform on $\mathbf{h}_1'$ to obtain

$$Z(\mathbf{m}_1, \mathbf{h}_2) = \sum_{\mathbf{h}_1' \in \mathbb{Z}^3/\mathbf{N}_1^T\mathbb{Z}^3} Z^*(\mathbf{h}_1', \mathbf{h}_2') e[-\mathbf{h}_1' \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)].$$

Putting $\mathbf{h}' = \mathbf{R}_g^T\mathbf{h}$ and using the symmetry of F in the form

$$F(\mathbf{h}') = F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{t}_g),$$

where

$$\begin{aligned} \mathbf{h} \cdot \mathbf{t}_g &= (\mathbf{h}_2^T + \mathbf{h}_1^T\mathbf{N}_2)(\mathbf{N}_2^{-1}\mathbf{N}_1^{-1})(\mathbf{t}_g^{(1)} + \mathbf{N}_1\mathbf{t}_g^{(2)}) \\ &\equiv \mathbf{h}_2 \cdot \mathbf{t}_g + \mathbf{h}_2 \cdot (\mathbf{N}_1^{-1}\mathbf{t}_g^{(1)}) \bmod 1 \end{aligned}$$

yields by a straightforward rearrangement

$$\begin{aligned} Z(\mathbf{m}_1, [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2) = e[-\{\mathbf{h}_2 \cdot \mathbf{t}_g + \boldsymbol{\eta}_1(g, \mathbf{h}_2) \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)\}] \\ \times Z\{S_g^{(1)}(\mathbf{m}_1), \mathbf{h}_2\}. \end{aligned}$$

This formula allows the transposition of intermediate results Z from an indexing by

$$(\text{all } \mathbf{m}_1) \times (\text{unique } \mathbf{h}_2)$$

to an indexing by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{h}_2).$$

We may then apply the twiddle factors to obtain

$$Y^*(\mathbf{m}_1, \mathbf{h}_2) = e[-\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)]Z(\mathbf{m}_1, \mathbf{h}_2)$$

and carry out the second transform on $\mathbf{h}_2$

$$Y(\mathbf{m}_1, \mathbf{m}_2) = \sum_{\mathbf{h}_2 \in \mathbb{Z}^3/\mathbf{N}_2^T\mathbb{Z}^3} Y^*(\mathbf{m}_1, \mathbf{h}_2) e[-\mathbf{h}_2 \cdot (\mathbf{N}_2^{-1}\mathbf{m}_2)].$$

The results, indexed by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{m}_2)$$

yield essentially an asymmetric unit of electron densities by the rearrangement

$$\not\rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) = Y(\mathbf{m}_1, \mathbf{m}_2).$$

The equivalence of the two transposition formulae up to the intervening twiddle factors is readily established, using the relation

$$\mathbf{h}_2 \cdot [\mathbf{N}_2^{-1}\boldsymbol{\mu}_2(g, \mathbf{m}_1)] = \boldsymbol{\eta}_1(g, \mathbf{h}_2) \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1) \bmod 1$$

which is itself a straightforward consequence of the identity

$$\mathbf{h} \cdot [\mathbf{N}^{-1}S_g(\mathbf{m})] = \mathbf{h} \cdot \mathbf{t}_g + (\mathbf{R}_g^T\mathbf{h}) \cdot (\mathbf{N}^{-1}\mathbf{m}).$$

To complete the characterization of the effect of symmetry on the Cooley–Tukey factorization, and of the economy of computation it allows, it remains to consider the possibility that some values of $\mathbf{m}_1$ may be invariant under some transformations $g \in G$ under the action $\mathbf{m}_1 \longmapsto S_g^{(1)}(\mathbf{m}_1)$.

Suppose that $\mathbf{m}_1$ has a nontrivial isotropy subgroup $G_{\mathbf{m}_1}$, and let $g \in G_{\mathbf{m}_1}$. Then each subarray $Y_{\mathbf{m}_1}$ defined by

$$Y_{\mathbf{m}_1}(\mathbf{m}_2) = Y(\mathbf{m}_1, \mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$$

satisfies the identity

$$Y_{\mathbf{m}_1}(\mathbf{m}_2) = Y_{S_g^{(1)}(\mathbf{m}_1)}[S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1)]$$
$$= Y_{\mathbf{m}_1}[S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1)]$$

so that the data for the transform on $\mathbf{m}_2$ have residual symmetry properties. In this case the identity satisfied by $\boldsymbol{\mu}_2$ simplifies to

$$\boldsymbol{\mu}_2(gg', \mathbf{m}_1) = S_g^{(2)}[\boldsymbol{\mu}_2(g', \mathbf{m}_1)] + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^3,$$

which shows that the mapping $g \longmapsto \boldsymbol{\mu}_2(g, \mathbf{m}_1)$ satisfies the Frobenius congruences (Section 1.3.4.2.2.3). Thus the internal symmetry of subarray $Y_{\mathbf{m}_1}$ with respect to the action of G on $\mathbf{m}_2$ is given by $G_{\mathbf{m}_1}$ acting on $\mathbb{Z}^3/\mathbf{N}_2\mathbb{Z}^3$ *via*

$$\mathbf{m}_2 \longmapsto S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^3.$$

The transform on $\mathbf{m}_2$ needs only be performed for one out of $[G : G_{\mathbf{m}_1}]$ distinct arrays $Y_{\mathbf{m}_1}$ (results for the others being obtainable by the transposition formula), and this transforms is $G_{\mathbf{m}_1}$-symmetric. In other words, the following cases occur:

(i) $G_{\mathbf{m}_1} = \{e\}$ — maximum saving in computation (by $|G|$); $\mathbf{m}_2$-transform has no symmetry.

(ii) $G_{\mathbf{m}_1} = G' < G$ — saving in computation by a factor of $[G : G']$; $\mathbf{m}_2$-transform is G'-symmetric.

(iii) $G_{\mathbf{m}_1} = G$ — no saving in computation; $\mathbf{m}_2$-transform is G-symmetric.

The symmetry properties of the $\mathbf{m}_2$-transform may themselves be exploited in a similar way if $\mathbf{N}_2$ can be factored as a product of smaller decimation matrices; otherwise, an appropriate symmetrized DFT routine may be provided, using for instance the idea of 'multiplexing/demultiplexing' (Section 1.3.4.3.5). We thus have a recursive *descent procedure*, in which the deeper stages of the recursion deal with transforms on *fewer points*, or of *lower symmetry* (usually both).

The same analysis applies to the $\mathbf{h}_1$-transforms on the subarrays $Z^*_{\mathbf{h}_2}$, and leads to a similar descent procedure.

In conclusion, crystallographic symmetry can be fully exploited to reduce the amount of computation to the minimum required to obtain the unique results from the unique data. No such analysis was so far available in cases where the asymmetric units in real and reciprocal space are not parallelepipeds. An example of this procedure will be given in Section 1.3.4.3.6.5.

1.3.4.3.4.2. Multidimensional Good factorization

This procedure was described in Section 1.3.3.3.2.2. The main difference with the Cooley–Tukey factorization is that if $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2 \ldots \mathbf{N}_{d-1}\mathbf{N}_d$, where the different factors are pairwise coprime, then the Chinese remainder theorem reindexing makes $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$ isomorphic to a direct sum.

$$\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3 \cong (\mathbb{Z}^3/\mathbf{N}_1\mathbb{Z}^3) \oplus \ldots \oplus (\mathbb{Z}^3/\mathbf{N}_d\mathbb{Z}^3),$$

where each p-primary piece is endowed with an induced $\mathbb{Z}G$-module structure by letting G operate in the usual way but with the corresponding modular arithmetic. The situation is thus more favourable than with the Cooley–Tukey method, since there is no interference between the factors (no 'carry'). In the terminology of Section 1.3.4.2.2.2, G acts *diagonally* on this direct sum, and results of a partial transform may be transposed by orbit exchange as in Section 1.3.4.3.4.1 but without the extra terms $\boldsymbol{\mu}$ or $\boldsymbol{\eta}$. The analysis of the symmetry properties of partial transforms also carries over, again without the extra terms. Further simplification occurs for all p-primary pieces with p other than 2 or 3, since all nonprimitive translations (including those associated to lattice centring) disappear modulo p.

Thus the cost of the CRT reindexing is compensated by the computational savings due to the absence of twiddle factors and of other phase shifts associated with nonprimitive translations and with geometric 'carries'.

Within each p-primary piece, however, higher powers of p may need to be split up by a Cooley–Tukey factorization, or carried out directly by a suitably adapted Winograd algorithm.

1.3.4.3.4.3. Crystallographic extension of the Rader/Winograd factorization

As was the case in the absence of symmetry, the two previous classes of algorithms can only factor the global transform into partial transforms on prime numbers of points, but cannot break the latter down any further. Rader's idea of using the action of the group of units $U(p)$ to obtain further factorization of a p-primary transform has been used in 'scalar' form by Auslander & Shenefelt (1987), Shenefelt (1988) and Auslander *et al.* (1988). It will be shown here that it can be adapted to the crystallographic case so as to take advantage also of the possible existence of n-fold cyclic symmetry elements ($n = 3, 4, 6$) in a two-dimensional transform (Bricogne & Tolimieri, 1990). This adaptation entails the use of certain rings of *algebraic* integers rather than ordinary integers, whose connection with the handling of cyclic symmetry will now be examined.

Let G be the group associated with a threefold axis of symmetry: $G = \{e, g, g^2\}$ with $g^3 = e$. In a standard trigonal basis, G has matrix representation

$$\mathbf{R}_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \mathbf{R}_g = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{R}_{g^2} = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

in real space,

$$\mathbf{R}^*_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \mathbf{R}^*_g = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{R}^*_{g^2} = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$$

in reciprocal space. Note that

$$\mathbf{R}^*_{g^2} = [\mathbf{R}^{-1}_{g^2}]^T = \mathbf{R}^T_g,$$

and that

$$\mathbf{R}^T_g = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}, \qquad \text{where } \mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

so that $\mathbf{R}_g$ and $\mathbf{R}^T_g$ are conjugate in the group of 2×2 unimodular integer matrices. The group ring $\mathbb{Z}G$ is commutative, and has the structure of the polynomial ring $\mathbb{Z}[X]$ with the single relation $X^2 + X + 1 = 0$ corresponding to the minimal polynomial of $\mathbf{R}_g$. In the terminology of Section 1.3.3.2.4, the ring structure of $\mathbb{Z}G$ is obtained from that of $\mathbb{Z}[X]$ by carrying out polynomial addition and multiplication modulo $X^2 + X + 1$, then replacing X by any generator of G. This type of construction forms the very basis of algebraic number theory [see Artin (1944, Section IIc) for an illustration of this viewpoint], and $\mathbb{Z}G$ as just defined is isomorphic to the ring $\mathbb{Z}[\omega]$ of algebraic integers of the form

$a + b\omega$ [$a, b \in \mathbb{Z}$, $\omega = \exp(2\pi i/3)$] under the identification $X \leftrightarrow \omega$. Addition in this ring is defined component-wise, while multiplication is defined by

$$(a_1 + b_1\omega) \times (a_2 + b_2\omega) = (a_1a_2 - b_1b_2) + [(a_1 - b_1)b_2 + b_1a_2]\omega.$$

In the case of a fourfold axis, $G = \{e, g, g^2, g^3\}$ with $g^4 = e$, and

$$\mathbf{R}_g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \mathbf{R}_g^*, \quad \text{with again } \mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}.$$

$\mathbb{Z}G$ is obtained from $\mathbb{Z}[X]$ by carrying out polynomial arithmetic modulo $X^2 + 1$. This identifies $\mathbb{Z}G$ with the ring $\mathbb{Z}[i]$ of Gaussian integers of the form $a + bi$, in which addition takes place component-wise while multiplication is defined by

$$(a_1 + b_1i) \times (a_2 + b_2i) = (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i.$$

In the case of a sixfold axis, $G = \{e, g, g^2, g^3, g^4, g^5\}$ with $g^6 = e$, and

$$\mathbf{R}_g = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{R}_g^* = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}.$$

$\mathbb{Z}G$ is isomorphic to $\mathbb{Z}[\omega]$ under the mapping $g \leftrightarrow 1 + \omega$ since $(1 + \omega)^6 = 1$.

Thus in all cases $\mathbb{Z}G \cong \mathbb{Z}[X]/P(X)$ where $P(X)$ is an irreducible quadratic polynomial with integer coefficients.

The actions of G on lattices in real and reciprocal space (Sections 1.3.4.2.2.4, 1.3.4.2.2.5) extend naturally to actions of $\mathbb{Z}G$ on $\mathbb{Z}^2$ in which an element $z = a + bg$ of $\mathbb{Z}G$ acts *via*

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \longmapsto z\mathbf{m} = (a\mathbf{I} + b\mathbf{R}_g)\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}$$

in real space, and *via*

$$\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \longmapsto z\mathbf{h} = (a\mathbf{I} + b\mathbf{R}_g^T)\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

in reciprocal space. These two actions are related by conjugation, since

$$(a\mathbf{I} + b\mathbf{R}_g^T) = \mathbf{J}^{-1}(a\mathbf{I} + b\mathbf{R}_g)\mathbf{J}$$

and the following identity (which is fundamental in the sequel) holds:

$$(z\mathbf{h}) \cdot \mathbf{m} = \mathbf{h} \cdot (z\mathbf{m}) \quad \text{for all } \mathbf{m}, \mathbf{h} \in \mathbb{Z}^2.$$

Let us now consider the calculation of a $p \times p$ two-dimensional DFT with n-fold cyclic symmetry ($n = 3, 4, 6$) for an odd prime $p \geq 5$. Denote $\mathbb{Z}/p\mathbb{Z}$ by $\mathbb{Z}_p$. Both the data and the results of the DFT are indexed by $\mathbb{Z}_p \times \mathbb{Z}_p$: hence the action of $\mathbb{Z}G$ on these indices is in fact an action of $\mathbb{Z}_pG$, the latter being obtained from $\mathbb{Z}G$ by carrying out all integer arithmetic in $\mathbb{Z}G$ modulo p. The algebraic structure of $\mathbb{Z}_pG$ combines the symmetry-carrying ring structure of $\mathbb{Z}G$ with the finite field structure of $\mathbb{Z}_p$ used in Section 1.3.3.2.3.1, and holds the key to a symmetry-adapted factorization of the DFT at hand.

The structure of $\mathbb{Z}_pG$ depends on whether $P(X)$ remains irreducible when considered as a polynomial over $\mathbb{Z}_p$. Thus two cases arise:

(1) $P(X)$ remains irreducible mod p, *i.e.* there is no nth root of unity in $\mathbb{Z}_p$;

(2) $P(X)$ factors as $(X - u)(X - v)$, *i.e.* there are nth roots of unity in $\mathbb{Z}_p$.

These two cases require different developments.

Case 1. $\mathbb{Z}_pG$ is a finite field with p^2 elements. There is essentially (*i.e.* up to isomorphism) only one such field, denoted $GF(p^2)$, and its group of units is a cyclic group with $p^2 - 1$ elements. If γ is a generator of this group of units, the input data $\rho_{\mathbf{m}}$ with $\mathbf{m} \neq \mathbf{0}$ may be reordered as

$$\mathbf{m}_0, \gamma\mathbf{m}_0, \gamma^2\mathbf{m}_0, \gamma^3\mathbf{m}_0, \ldots, \gamma^{p^2-2}\mathbf{m}_0$$

by the *real-space action* of γ; while the results $F_{\mathbf{h}}$ with $\mathbf{h} \neq \mathbf{0}$ may be reordered as

$$\mathbf{h}_0, \gamma\mathbf{h}_0, \gamma^2\mathbf{h}_0, \gamma^3\mathbf{h}_0, \ldots, \gamma^{p^2-2}\mathbf{h}_0$$

by the *reciprocal-space action* of γ, where $\mathbf{m}_0$ and $\mathbf{h}_0$ are arbitrary nonzero indices.

The core $\mathbf{C}_{p\times p}$ of the DFT matrix, defined by

$$\mathbf{F}_{p\times p} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & & & \\ \vdots & & \mathbf{C}_{p\times p} & \\ 1 & & & \end{pmatrix},$$

will then have a skew-circulant structure (Section 1.3.3.2.3.1) since

$$(\mathbf{C}_{p\times p})_{jk} = e\left[\frac{(\gamma^j\mathbf{h}_0) \cdot (\gamma^k\mathbf{m}_0)}{p}\right] = e\left[\frac{\mathbf{h}_0 \cdot (\gamma^{j+k}\mathbf{m}_0)}{p}\right]$$

depends only on $j + k$. Multiplication by $\mathbf{C}_{p\times p}$ may then be turned into a cyclic convolution of length $p^2 - 1$, which may be factored by two DFTs (Section 1.3.3.2.3.1) or by Winograd's techniques (Section 1.3.3.2.4). The latter factorization is always favourable, as it is easily shown that $p^2 - 1$ is divisible by 24 for any odd prime $p \geq 5$. This procedure is applicable even if no symmetry is present in the data.

Assume now that cyclic symmetry of order $n = 3$, 4 or 6 *is* present. Since n divides 24 hence divides $p^2 - 1$, the generator g of this symmetry is representable as $\gamma^{(p^2-1)/n}$ for a suitable generator γ of the group of units. The reordered data will then be $(p^2 - 1)/n$-periodic rather than simply $(p^2 - 1)$-periodic; hence the reindexed results will be *n-decimated* (Section 1.3.2.7.2), and the $(p^2 - 1)/n$ nonzero results can be calculated by applying the DFT to the $(p^2 - 1)/n$ unique input data. In this way, the n-fold symmetry can be used in full to calculate the core contributions from the unique data to the unique results by a DFT of length $(p^2 - 1)/n$.

It is a simple matter to incorporate nonprimitive translations into this scheme. For example, when going from structure factors to electron densities, reordered data items separated by $(p^2 - 1)/n$ are not equal but differ by a phase shift proportional to their index mod p, whose effect is simply to shift the origin of the n-decimated transformed sequence. The same economy of

computation can therefore be achieved as in the purely cyclic case.

Dihedral symmetry elements, which map g to g^{-1} (Section 1.3.4.2.2.3), induce extra one-dimensional symmetries of order 2 in the reordered data which can also be fully exploited to reduce computation.

Case 2. If $p \geq 5$, it can be shown that the two roots u and v are always distinct. Then, by the Chinese remainder theorem (CRT) for polynomials (Section 1.3.3.2.4) we have a ring isomorphism

$$\mathbb{Z}_p[X]/P(X) \cong \{\mathbb{Z}_p[X]/(X-u)\} \times \{\mathbb{Z}_p[X]/(X-v)\}$$

defined by sending a polynomial $Q(X)$ from the left-hand-side ring to its two residue classes modulo $X-u$ and $X-v$, respectively. Since the latter are simply the constants $Q(u)$ and $Q(v)$, the CRT reindexing has the particularly simple form

$$a + bX \longmapsto (a+bu, a+bv) = (\alpha, \beta)$$

or equivalently

$$\begin{pmatrix} a \\ b \end{pmatrix} \longmapsto \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{M}\begin{pmatrix} a \\ b \end{pmatrix} \bmod p, \quad \text{with } \mathbf{M} = \begin{pmatrix} 1 & u \\ 1 & v \end{pmatrix}.$$

The CRT reconstruction formula similarly simplifies to

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \longmapsto \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{M}^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \bmod p,$$
$$\text{with } \mathbf{M}^{-1} = \frac{1}{v-u}\begin{pmatrix} v & -u \\ -1 & 1 \end{pmatrix}.$$

The use of the CRT therefore amounts to the *simultaneous diagonalization* (by $\mathbf{M}$) of all the matrices representing the elements of $\mathbb{Z}_p G$ in the basis $(1, X)$.

A first consequence of this diagonalization is that the internal structure of $\mathbb{Z}_p G$ becomes clearly visible. Indeed, $\mathbb{Z}_p G$ is mapped isomorphically to a direct product of two copies of $\mathbb{Z}_p$, in which arithmetic is carried out *component-wise* between eigenvalues α and β. Thus if

$$z = a + bX \overset{\text{CRT}}{\longleftrightarrow} (\alpha, \beta),$$
$$z' = a' + b'X \overset{\text{CRT}}{\longleftrightarrow} (\alpha', \beta'),$$

then

$$z + z' \overset{\text{CRT}}{\longleftrightarrow} (\alpha + \alpha', \beta + \beta'),$$
$$zz' \overset{\text{CRT}}{\longleftrightarrow} (\alpha\alpha', \beta\beta').$$

Taking in particular

$$z \overset{\text{CRT}}{\longleftrightarrow} (\alpha, 0) \neq (0, 0),$$
$$z' \overset{\text{CRT}}{\longleftrightarrow} (0, \beta) \neq (0, 0),$$

we have $zz' = 0$, so that $\mathbb{Z}_p G$ contains zero divisors; therefore $\mathbb{Z}_p G$ *is not a field.* On the other hand, if $z \overset{\text{CRT}}{\longleftrightarrow} (\alpha, \beta)$ with $\alpha \neq 0$ and $\beta \neq 0$, then α and β belong to the group of units $U(p)$ (Section 1.3.3.2.3.1) and hence have inverses α^{-1} and β^{-1}; it follows that z is a unit in $\mathbb{Z}_p G$, with inverse $z^{-1} \overset{\text{CRT}}{\longleftrightarrow} (\alpha^{-1}, \beta^{-1})$. Therefore, $\mathbb{Z}_p G$ consists of four distinct pieces:

$$0 \overset{\text{CRT}}{\longleftrightarrow} \{(0,0)\},$$
$$D_1 \overset{\text{CRT}}{\longleftrightarrow} \{(\alpha, 0) | \alpha \in U(p)\} \cong U(p),$$
$$D_2 \overset{\text{CRT}}{\longleftrightarrow} \{(0, \beta) | \beta \in U(p)\} \cong U(p),$$
$$U \overset{\text{CRT}}{\longleftrightarrow} \{(\alpha, \beta) | \alpha \in U(p), \beta \in U(p)\} \cong U(p) \times U(p).$$

A second consequence of this diagonalization is that the actions of $\mathbb{Z}_p G$ on indices $\mathbf{m}$ and $\mathbf{h}$ can themselves be brought to diagonal form by basis changes:

$$\mathbf{m} \longmapsto (a\mathbf{I} + b\mathbf{R}_g)\mathbf{m}$$
$$\text{becomes } \boldsymbol{\mu} \longmapsto \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\boldsymbol{\mu} \quad \text{with } \boldsymbol{\mu} = \mathbf{M}\mathbf{m},$$
$$\mathbf{h} \longmapsto (a\mathbf{I} + b\mathbf{R}_g^T)\mathbf{h}$$
$$\text{becomes } \boldsymbol{\eta} \longmapsto \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\boldsymbol{\eta} \quad \text{with } \boldsymbol{\eta} = \mathbf{M}\mathbf{J}\mathbf{h}.$$

Thus the sets of indices $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ can be split into four pieces as $\mathbb{Z}_p G$ itself, according as these indices have none, one or two of their coordinates in $U(p)$. These pieces will be labelled by the same symbols – 0, D_1, D_2 and U – as those of $\mathbb{Z}_p G$.

The scalar product $\mathbf{h} \cdot \mathbf{m}$ may be written in terms of $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ as

$$\mathbf{h} \cdot \mathbf{m} = [\boldsymbol{\eta} \cdot ((\mathbf{M}^{-1})^T \mathbf{J} \mathbf{M}^{-1})\boldsymbol{\mu}],$$

and an elementary calculation shows that the matrix $\boldsymbol{\Delta} = (\mathbf{M}^{-1})^T \mathbf{J} \mathbf{M}^{-1}$ is *diagonal* by virtue of the relation

$$uv = \text{constant term in } P(X) = 1.$$

Therefore, $\mathbf{h} \cdot \mathbf{m} = 0$ if $\mathbf{h} \in D_1$ and $\boldsymbol{\mu} \in D_2$ or *vice versa.*

We are now in a position to rearrange the DFT matrix $\mathbf{F}_{p \times p}$. Clearly, the structure of $\mathbf{F}_{p \times p}$ is more complex than in case 1, as there are three types of 'core' matrices:

type 1: $D \times D$ (with $D = D_1$ or D_2);
type 2: $D \times U$ or $U \times D$;
type 3: $U \times U$.

(Submatrices of type $D_1 \times D_2$ and $D_2 \times D_1$ have all their elements equal to 1 by the previous remark.)

Let γ be a generator of $U(p)$. We may reorder the elements in D_1, D_2 and U – and hence the data and results indexed by these elements – according to powers of γ. This requires one exponent in each of D_1 and D_2, and two exponents in U. For instance, in the $\mathbf{h}$-index space:

$$D_1 = \left\{ \begin{pmatrix} \gamma & 0 \\ 0 & 0 \end{pmatrix}^j \begin{pmatrix} \eta_1 \\ 0 \end{pmatrix}_0 \Big| j = 1, \ldots, p-1 \right\}$$

$$D_2 = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}^j \begin{pmatrix} 0 \\ \eta_2 \end{pmatrix}_0 \Big| j = 1, \ldots, p-1 \right\}$$

$$U = \left\{ \begin{pmatrix} \gamma & 0 \\ 0 & 1 \end{pmatrix}^{j_1} \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}^{j_2} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}_0 \Big| j_1 = 1, \ldots, p-1; \; j_2 = 1, \ldots, p-1 \right\}$$

and similarly for the $\boldsymbol{\mu}$ index.

Since the diagonal matrix $\boldsymbol{\Delta}$ commutes with all the matrices representing the action of γ, this rearrangement will induce skew-circulant structures in all the core matrices. The corresponding cyclic convolutions may be carried out by Rader's method, *i.e.* by diagonalizing them by means of two $(p-1)$-point one-dimensional DFTs in the $D \times D$ pieces and of two $(p-1) \times (p-1)$-point two-dimensional DFTs in the $U \times U$ piece (the $U \times D$ and $D \times U$ pieces involve extra section and projection operations).

In the absence of symmetry, no computational saving is achieved, since the same reordering could have been applied to the initial $\mathbb{Z}_p \times \mathbb{Z}_p$ indexing, without the CRT reindexing.

In the presence of n-fold cyclic symmetry, however, the rearranged $\mathbf{F}_{p\times p}$ lends itself to an n-fold reduction in size. The basic fact is that whenever case 2 occurs, $p-1$ is divisible by n (*i.e.* $p-1$ is divisible by 6 when $n = 3$ or 6, and by 4 when $n = 4$), say $p - 1 = nq$. If g is a generator of the cyclic symmetry, the generator γ of $U(p)$ may be chosen in such a way that $g = \gamma^q$. The action of g is then to increment the j index in D_1 and D_2 by q, and the (j_1, j_2) index in U by (q, q). Since the data items whose indices are related in this way have identical values, the DFTs used to diagonalize the Rader cyclic convolutions will operate on *periodized data*, hence yield *decimated results*; and the nonzero results will be obtained from the unique data by DFTs n times smaller than their counterparts in the absence of symmetry.

A more thorough analysis is needed to obtain a Winograd factorization into the normal from **CBA** in the presence of symmetry (see Bricogne & Tolimieri, 1990).

Nonprimitive translations and dihedral symmetry may also be accommodated within this framework, as in case 1.

This reindexing by means of algebraic integers yields larger orbits, hence more efficient algorithms, than that of Auslander *et al.* (1988) which only uses ordinary integers acting by scalar dilation.

1.3.4.3.5. Treatment of conjugate and parity-related symmetry properties

Most crystallographic Fourier syntheses are real-valued and originate from Hermitian-symmetric collections of Fourier coefficients. Hermitian symmetry is closely related to the action of a centre of inversion in reciprocal space, and thus interacts strongly with all other genuinely crystallographic symmetry elements of order 2. All these symmetry properties are best treated by factoring by 2 and reducing the computation of the initial transform to that of a collection of smaller transforms with less symmetry or none at all.

1.3.4.3.5.1. Hermitian-symmetric or real-valued transforms

The computation of a DFT with Hermitian-symmetric or real-valued data can be carried out at a cost of half that of an ordinary transform, essentially by 'multiplexing' pairs of special partial transforms into general complex transforms, and then 'demultiplexing' the results on the basis of their symmetry properties. The treatment given below is for general dimension n; a subset of cases for $n = 1$ was treated by Ten Eyck (1973).

(*a*) *Underlying group action*

Hermitian symmetry is not a geometric symmetry, but it is defined in terms of the action in reciprocal space of point group $G = \bar{1}$, *i.e.* $G = \{e, -e\}$, where e acts as $\mathbf{I}$ (the $n \times n$ identity matrix) and $-e$ acts as $-\mathbf{I}$.

This group action on $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ with $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2$ will now be characterized by the calculation of the cocycle $\boldsymbol{\eta}_1$ (Section 1.3.4.3.4.1) under the assumption that $\mathbf{N}_1$ and $\mathbf{N}_2$ are both *diagonal*. For this purpose it is convenient to associate to any integer vector

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

in $\mathbb{Z}^n$ the vector $\boldsymbol{\zeta}(\mathbf{v})$ whose jth component is

$$\begin{cases} 0 \text{ if } & v_j = 0 \\ 1 \text{ if } & v_j \neq 0. \end{cases}$$

Let $\mathbf{m} = \mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2$, and hence $\mathbf{h} = \mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1$. Then

$$-\mathbf{h}_2 \bmod \mathbf{N}\mathbb{Z}^n = \mathbf{N}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2,$$
$$-\mathbf{h}_2 \bmod \mathbf{N}_2\mathbb{Z}^n = \mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2,$$

hence

$$\begin{aligned} \boldsymbol{\eta}_1(-e, \mathbf{h}_2) &= \mathbf{N}_2^{-1}\{[\mathbf{N}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2] - [\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2]\} \bmod \mathbf{N}_1\mathbb{Z}^n \\ &= -\boldsymbol{\zeta}(\mathbf{h}_2) \bmod \mathbf{N}_1\mathbb{Z}^n. \end{aligned}$$

Therefore $-e$ acts by

$$(\mathbf{h}_2, \mathbf{h}_1) \longmapsto [\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2, \mathbf{N}_1\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1 - \boldsymbol{\zeta}(\mathbf{h}_2)].$$

Hermitian symmetry is traditionally dealt with by factoring by 2, *i.e.* by assuming $\mathbf{N} = 2\mathbf{M}$. If $\mathbf{N}_2 = 2\mathbf{I}$, then each $\mathbf{h}_2$ is invariant under G, so that each partial vector $\mathbf{Z}^*_{\mathbf{h}_2}$ (Section 1.3.4.3.4.1) inherits the symmetry internally, with a 'modulation' by $\boldsymbol{\eta}_1(g, \mathbf{h}_2)$. The 'multiplexing–demultiplexing' technique provides an efficient treatment of this singular case.

(*b*) *Calculation of structure factors*

The computation may be summarized as follows:

$$\mathcal{P} \overset{\mathbf{dec}(\mathbf{N}_1)}{\longmapsto} \mathbf{Y} \overset{\bar{F}(\mathbf{N}_2)}{\longmapsto} \mathbf{Y}^* \overset{\text{TW}}{\longmapsto} \mathbf{Z} \overset{\bar{F}(\mathbf{N}_1)}{\longmapsto} \mathbf{Z}^* \overset{\mathbf{rev}(\mathbf{N}_2)}{\longmapsto} \mathbf{F}$$

where $\mathbf{dec}(\mathbf{N}_1)$ is the initial decimation given by $Y_{\mathbf{m}_1}(\mathbf{m}_2) = \mathcal{P}(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$, TW is the transposition and twiddle-factor stage, and $\mathbf{rev}(\mathbf{N}_2)$ is the final unscrambling by coset reversal given by $F(\mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1) = \mathbf{Z}^*_{\mathbf{h}_2}(\mathbf{h}_1)$.

(i) *Decimation in time* ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

The decimated vectors $\mathbf{Y}_{\mathbf{m}_1}$ are real and hence have Hermitian transforms $\mathbf{Y}^*_{\mathbf{m}_1}$. The 2^n values of $\mathbf{m}_1$ may be grouped into 2^{n-1} pairs $(\mathbf{m}_1', \mathbf{m}_1'')$ and the vectors corresponding to each pair may be multiplexed into a general complex vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}_1'} + i\mathbf{Y}_{\mathbf{m}_1''}.$$

The transform $\mathbf{Y}^* = \bar{F}(\mathbf{M})[\mathbf{Y}]$ can then be resolved into the separate transforms $\mathbf{Y}^*_{\mathbf{m}_1'}$ and $\mathbf{Y}^*_{\mathbf{m}_1''}$ by using the Hermitian symmetry of the latter, which yields the demultiplexing formulae

$$\begin{aligned} Y^*_{\mathbf{m}_1'}(\mathbf{h}_2) + iY^*_{\mathbf{m}_1''}(\mathbf{h}_2) &= Y^*(\mathbf{h}_2) \\ \overline{Y^*_{\mathbf{m}_1'}(\mathbf{h}_2)} + i\overline{Y^*_{\mathbf{m}_1''}(\mathbf{h}_2)} &= Y^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2]. \end{aligned}$$

The number of partial transforms $\bar{F}(\mathbf{M})$ is thus reduced from 2^n to 2^{n-1}. Once this separation has been achieved, the remaining steps need only be carried out for a unique half of the values of $\mathbf{h}_2$.

(ii) *Decimation in frequency* ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

Since $\mathbf{h}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{h}_2 = \mathbf{h}_2$ and $\boldsymbol{\zeta}(\mathbf{h}_2) = \mathbf{h}_2 \bmod 2\mathbb{Z}^n$. The vectors of decimated and scrambled results $\mathbf{Z}^*_{\mathbf{h}_2}$ then obey the symmetry relations

$$Z^*_{\mathbf{h}_2}(\mathbf{h}_1 - \mathbf{h}_2) = \overline{Z^*_{\mathbf{h}_2}[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1]}$$

which can be used to halve the number of $\bar{F}(\mathbf{M})$ necessary to compute them, as follows.

Having formed the vectors $\mathbf{Z}_{\mathbf{h}_2}$ given by

$$Z_{\mathbf{h}_2}(\mathbf{m}_1) = \left[\sum_{\mathbf{m}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n} \frac{(-1)^{\mathbf{h}_2\cdot\mathbf{m}_2}}{2^n} \rho(\mathbf{m}_1 + \mathbf{M}\mathbf{m}_2)\right] e[\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)],$$

we may group the 2^n values of $\mathbf{h}_2$ into 2^{n-1} pairs $(\mathbf{h}_2', \mathbf{h}_2'')$ and for each pair form the multiplexed vector:

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}_2'} + i\mathbf{Z}_{\mathbf{h}_2''}.$$

After calculating the 2^{n-1} transforms $\mathbf{Z}^* = \bar{F}(\mathbf{M})[\mathbf{Z}]$, the 2^n individual transforms $\mathbf{Z}^*_{\mathbf{h}_2'}$ and $\mathbf{Z}^*_{\mathbf{h}_2''}$ can be separated by using for each pair the demultiplexing formulae

$$\begin{aligned} Z^*_{\mathbf{h}_2'}(\mathbf{h}_1) + iZ^*_{\mathbf{h}_2''}(\mathbf{h}_1) &= Z^*(\mathbf{h}_1) \\ Z^*_{\mathbf{h}_2'}(\mathbf{h}_1 - \mathbf{h}_2') + iZ^*_{\mathbf{h}_2''}(\mathbf{h}_1 - \mathbf{h}_2'') &= \overline{Z^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1]} \end{aligned}$$

which can be solved recursively. If all pairs are chosen so that they differ only in the jth coordinate $(\mathbf{h}_2)_j$, the recursion is along $(\mathbf{h}_1)_j$ and can be initiated by introducing the (real) values of $Z^*_{\mathbf{h}_2'}$ and $Z^*_{\mathbf{h}_2''}$ at $(\mathbf{h}_1)_j = 0$ and $(\mathbf{h}_1)_j = M_j$, accumulated *e.g.* while forming $\mathbf{Z}$ for that pair. Only points with $(\mathbf{h}_1)_j$ going from 0 to $\frac{1}{2}M_j$ need be resolved, and they contain the unique half of the Hermitian-symmetric transform $\mathbf{F}$.

(*c*) *Calculation of electron densities*

The computation may be summarized as follows:

$$\mathbf{F} \overset{\mathbf{scr}(\mathbf{N}_2)}{\longmapsto} \mathbf{Z}^* \overset{F(\mathbf{N}_1)}{\longmapsto} \mathbf{Z} \overset{\mathrm{TW}}{\longmapsto} \mathbf{Y}^* \overset{F(\mathbf{N}_2)}{\longmapsto} \mathbf{Y} \overset{\mathbf{nat}(\mathbf{N}_1)}{\longmapsto} \rho$$

where $\mathbf{scr}(\mathbf{N}_2)$ is the decimation with coset reversal given by $\mathbf{Z}^*_{\mathbf{h}_2}(\mathbf{h}_1) = F(\mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1)$, TW is the transposition and twiddle-factor stage, and $\mathbf{nat}(\mathbf{N}_1)$ is the recovery in natural order given by $\rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) = Y_{\mathbf{m}_1}(\mathbf{m}_2)$.

(i) *Decimation in time* ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

The last transformation $F(2\mathbf{I})$ has a real-valued matrix, and the final result ρ is real-valued. It follows that the vectors $\mathbf{Y}^*_{\mathbf{m}_1}$ of intermediate results after the twiddle-factor stage are real-valued, hence lend themselves to multiplexing along the real and imaginary components of half as many general complex vectors.

Let the 2^n initial vectors $\mathbf{Z}^*_{\mathbf{h}_2}$ be multiplexed into 2^{n-1} vectors

$$\mathbf{Z}^* = \mathbf{Z}^*_{\mathbf{h}_2'} + i\mathbf{Z}^*_{\mathbf{h}_2''}$$

[one for each pair $(\mathbf{h}_2', \mathbf{h}_2'')$], each of which yields by $F(\mathbf{M})$ a vector

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}_2'} + i\mathbf{Z}_{\mathbf{h}_2''}.$$

The real-valuedness of the $\mathbf{Y}^*_{\mathbf{m}_1}$ may be used to recover the separate result vectors for $\mathbf{h}_2'$ and $\mathbf{h}_2''$. For this purpose, introduce the abbreviated notation

$$\begin{aligned} e[-\mathbf{h}_2' \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] &= (c' + is')(\mathbf{m}_1) \\ e[-\mathbf{h}_2'' \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] &= (c'' + is'')(\mathbf{m}_1) \\ R_{\mathbf{h}_2}(\mathbf{m}_1) &= Y^*_{\mathbf{m}_1}(\mathbf{h}_2) \\ \mathbf{R}' = \mathbf{R}_{\mathbf{h}_2'}, &\quad \mathbf{R}'' = \mathbf{R}_{\mathbf{h}_2''}. \end{aligned}$$

Then we may write

$$\begin{aligned} \mathbf{Z} &= (c' + is')\mathbf{R}' + i(c'' + is'')\mathbf{R}'' \\ &= (c'\mathbf{R}' + s''\mathbf{R}'') + i(-s'\mathbf{R}' + c''\mathbf{R}'') \end{aligned}$$

or, equivalently, for each $\mathbf{m}_1$,

$$\begin{pmatrix} \mathscr{R}e\, Z \\ \mathscr{I}m\, Z \end{pmatrix} = \begin{pmatrix} c' & s'' \\ -s' & c'' \end{pmatrix} \begin{pmatrix} R' \\ R'' \end{pmatrix}.$$

Therefore $\mathbf{R}'$ and $\mathbf{R}''$ may be retrieved from $\mathbf{Z}$ by the 'demultiplexing' formula:

$$\begin{pmatrix} R' \\ R'' \end{pmatrix} = \frac{1}{c'c'' + s's''} \begin{pmatrix} c'' & -s'' \\ s' & c' \end{pmatrix} \begin{pmatrix} \mathscr{R}e\, Z \\ \mathscr{I}m\, Z \end{pmatrix}$$

which is valid at all points $\mathbf{m}_1$ where $c'c'' + s's'' \neq 0$, *i.e.* where

$$\cos[2\pi(\mathbf{h}_2' - \mathbf{h}_2'') \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] \neq 0.$$

Demultiplexing fails when

$$(\mathbf{h}_2' - \mathbf{h}_2'') \cdot (\mathbf{N}^{-1}\mathbf{m}_1) = \tfrac{1}{2} \bmod 1.$$

If the pairs $(\mathbf{h}_2', \mathbf{h}_2'')$ are chosen so that their members differ only in one coordinate (the jth, say), then the exceptional points are at $(\mathbf{m}_1)_j = \frac{1}{2}M_j$ and the missing transform values are easily obtained *e.g.* by accumulation while forming $\mathbf{Z}^*$.

The final stage of the calculation is then

$$\rho(\mathbf{m}_1 + \mathbf{M}\mathbf{m}_2) = \sum_{\mathbf{h}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n} (-1)^{\mathbf{h}_2\cdot\mathbf{m}_2} R_{\mathbf{h}_2}(\mathbf{m}_1).$$

(ii) *Decimation in frequency* ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

The last transformation $F(\mathbf{M})$ gives the real-valued results ρ, therefore the vectors $\mathbf{Y}^*_{\mathbf{m}_1}$ after the twiddle-factor stage each have Hermitian symmetry.

A first consequence is that the intermediate vectors $\mathbf{Z}_{\mathbf{h}_2}$ need only be computed for the unique half of the values of $\mathbf{h}_2$, the other half being related by the Hermitian symmetry of $\mathbf{Y}^*_{\mathbf{m}_1}$.

A second consequence is that the 2^n vectors $\mathbf{Y}^*_{\mathbf{m}_1}$ may be condensed into 2^{n-1} general complex vectors

$$\mathbf{Y}^* = \mathbf{Y}^*_{\mathbf{m}'_1} + i\mathbf{Y}^*_{\mathbf{m}''_1}$$

[one for each pair $(\mathbf{m}'_1, \mathbf{m}''_1)$] to which a general complex $F(\mathbf{M})$ may be applied to yield

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + i\mathbf{Y}_{\mathbf{m}''_1}$$

with $\mathbf{Y}_{\mathbf{m}'_1}$ and $\mathbf{Y}_{\mathbf{m}''_1}$ *real-valued*. The final results can therefore be retrieved by the particularly simple demultiplexing formulae:

$$\rho(\mathbf{m}'_1 + 2\mathbf{m}_2) = \mathscr{R}e\, Y(\mathbf{m}_2),$$
$$\rho(\mathbf{m}''_1 + 2\mathbf{m}_2) = \mathscr{I}m\, Y(\mathbf{m}_2).$$

1.3.4.3.5.2. Hermitian-antisymmetric or pure imaginary transforms

A vector $\mathbf{X} = \{X(\mathbf{k}) | \mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n\}$ is said to be Hermitian-antisymmetric if

$$X(\mathbf{k}) = -\overline{X(-\mathbf{k})} \text{ for all } \mathbf{k}.$$

Its transform $\mathbf{X}^*$ then satisfies

$$X^*(\mathbf{k}^*) = -\overline{X^*(\mathbf{k}^*)} \text{ for all } \mathbf{k}^*,$$

i.e. is purely imaginary.

If $\mathbf{X}$ is Hermitian-antisymmetric, then $\mathbf{F} = \pm i\mathbf{X}$ is Hermitian-symmetric, with $\rho = \pm i\mathbf{X}^*$ real-valued. The treatment of Section 1.3.4.3.5.1 may therefore be adapted, with trivial factors of i or -1, or used as such in conjunction with changes of variable by multiplication by $\pm i$.

1.3.4.3.5.3. Complex symmetric and antisymmetric transforms

The matrix $-\mathbf{I}$ is its own contragredient, and hence (Section 1.3.2.4.2.2) the transform of a symmetric (respectively antisymmetric) function is symmetric (respectively antisymmetric). In this case the group $G = \{e, -e\}$ acts in both real and reciprocal space as $\{\mathbf{I}, -\mathbf{I}\}$. If $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2$ with both factors diagonal, then $-e$ acts by

$$(\mathbf{m}_1, \mathbf{m}_2) \longmapsto [\mathbf{N}_1\boldsymbol{\zeta}(\mathbf{m}_1) - \mathbf{m}_1, \mathbf{N}_2\boldsymbol{\zeta}(\mathbf{m}_2) - \mathbf{m}_2 - \boldsymbol{\zeta}(\mathbf{m}_1)],$$
$$(\mathbf{h}_2, \mathbf{h}_1) \longmapsto [\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2, \mathbf{N}_1\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1 - \boldsymbol{\zeta}(\mathbf{h}_2)],$$

i.e.

$$\boldsymbol{\mu}_2(-e, \mathbf{m}_1) = -\boldsymbol{\zeta}(\mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^n,$$
$$\boldsymbol{\eta}_1(-e, \mathbf{h}_2) = -\boldsymbol{\zeta}(\mathbf{h}_2) \bmod \mathbf{N}_1\mathbb{Z}^n.$$

The symmetry or antisymmetry properties of $\mathbf{X}$ may be written

$$X(-\mathbf{m}) = -\varepsilon X(\mathbf{m}) \text{ for all } \mathbf{m},$$

with $\varepsilon = +1$ for symmetry and $\varepsilon = -1$ for antisymmetry.

The computation will be summarized as

$$\mathbf{X} \overset{\mathbf{dec}(\mathbf{N}_1)}{\longmapsto} \mathbf{Y} \overset{\tilde{F}(\mathbf{N}_2)}{\longmapsto} \mathbf{Y}^* \overset{\mathrm{TW}}{\longmapsto} \mathbf{Z} \overset{\tilde{F}(\mathbf{N}_1)}{\longmapsto} \mathbf{Z}^* \overset{\mathbf{rev}(\mathbf{N}_2)}{\longmapsto} \mathbf{X}^*$$

with the same indexing as that used for structure-factor calculation. In both cases it will be shown that a transform $F(\mathbf{N})$ with $\mathbf{N} = \mathbf{2M}$ and $\mathbf{M}$ diagonal can be computed using only 2^{n-1} partial transforms $F(\mathbf{M})$ instead of 2^n.

(i) *Decimation in time* ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

Since $\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{m}_1 = \mathbf{m}_1$ and $\boldsymbol{\zeta}(\mathbf{m}_1) = \mathbf{m}_1 \bmod 2\mathbb{Z}^n$, so that the symmetry relations for each parity class of data $\mathbf{Y}_{\mathbf{m}_1}$ read

$$Y_{\mathbf{m}_1}[\mathbf{M}\boldsymbol{\zeta}(\mathbf{m}_2) - \mathbf{m}_2 - \mathbf{m}_1] = \varepsilon Y_{\mathbf{m}_1}(\mathbf{m}_2)$$

or equivalently

$$\tau_{\mathbf{m}_1}\mathbf{Y}_{\mathbf{m}_1} = \varepsilon\check{\mathbf{Y}}_{\mathbf{m}_1}.$$

Transforming by $F(\mathbf{M})$, this relation becomes

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}_1)]\mathbf{Y}^*_{\mathbf{m}_1} = \varepsilon\mathbf{Y}^*_{\mathbf{m}_1}.$$

Each parity class thus obeys a different symmetry relation, so that we may multiplex them in pairs by forming for each pair $(\mathbf{m}'_1, \mathbf{m}''_1)$ the vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + \mathbf{Y}_{\mathbf{m}''_1}.$$

Putting

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}'_1)] = (c' + is')(\mathbf{h}_2)$$
$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}''_1)] = (c'' + is'')(\mathbf{h}_2)$$

we then have the demultiplexing relations for each $\mathbf{h}_2$:

$$Y^*_{\mathbf{m}'_1}(\mathbf{h}_2) + Y^*_{\mathbf{m}''_1}(\mathbf{h}_2) = Y^*(\mathbf{h}_2)$$
$$(c' + is')(\mathbf{h}_2)Y^*_{\mathbf{m}'_1}(\mathbf{h}_2) + (c'' + is'')(\mathbf{h}_2)Y^*_{\mathbf{m}''_1}(\mathbf{h}_2)$$
$$= \varepsilon Y^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2]$$

which can be solved recursively. Transform values at the exceptional points $\mathbf{h}_2$ where demultiplexing fails (*i.e.* where $c' + is' = c'' + is''$) can be accumulated while forming $\mathbf{Y}$.

Only the unique half of the values of $\mathbf{h}_2$ need to be considered at the demultiplexing stage and at the subsequent TW and $F(2\mathbf{I})$ stages.

(ii) *Decimation in frequency* ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

The vectors of final results $\mathbf{Z}^*_{\mathbf{h}_2}$ for each parity class $\mathbf{h}_2$ obey the symmetry relations

$$\tau_{\mathbf{h}_2}\mathbf{Z}^*_{\mathbf{h}_2} = \varepsilon\check{\mathbf{Z}}^*_{\mathbf{h}_2},$$

which are different for each $\mathbf{h}_2$. The vectors $\mathbf{Z}_{\mathbf{h}_2}$ of intermediate results after the twiddle-factor stage may then be multiplexed in pairs as

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}_2'} + \mathbf{Z}_{\mathbf{h}_2''}.$$

After transforming by $F(\mathbf{M})$, the results $\mathbf{Z}^*$ may be demultiplexed by using the relations

$$\begin{aligned} Z^*_{\mathbf{h}_2'}(\mathbf{h}_1) \quad + Z^*_{\mathbf{h}_2''}(\mathbf{h}_1) \qquad &= Z^*(\mathbf{h}_1) \\ Z^*_{\mathbf{h}_2'}(\mathbf{h}_1 - \mathbf{h}_2') + Z^*_{\mathbf{h}_2''}(\mathbf{h}_1 - \mathbf{h}_2'') &= \varepsilon Z^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1] \end{aligned}$$

which can be solved recursively as in Section 1.3.4.3.5.1(*b*)(ii).

1.3.4.3.5.4. Real symmetric transforms

Conjugate symmetric (Section 1.3.2.4.2.3) implies that if the data $\mathbf{X}$ are real and symmetric [*i.e.* $X(\mathbf{k}) = \overline{X(\mathbf{k})}$ and $X(-\mathbf{k}) = X(\mathbf{k})$], then so are the results $\mathbf{X}^*$. Thus if $\mathcal{P}$ contains a centre of symmetry, $\mathbf{F}$ is real symmetric. There is no distinction (other than notation) between structure-factor and electron-density calculation; the algorithms will be described in terms of the former. It will be shown that if $\mathbf{N} = 2\mathbf{M}$, a real symmetric transform can be computed with only 2^{n-2} partial transforms $F(\mathbf{M})$ instead of 2^n.

(i) *Decimation in time* ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

Since $\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{m}_1 = \mathbf{m}_1$ and $\boldsymbol{\zeta}(\mathbf{m}_1) = \mathbf{m}_1 \bmod 2\mathbb{Z}^n$. The decimated vectors $\mathbf{Y}_{\mathbf{m}_1}$ are not only real, but have an internal symmetry expressed by

$$\mathbf{Y}_{\mathbf{m}_1}[\mathbf{M}\boldsymbol{\zeta}(\mathbf{m}_2) - \mathbf{m}_2 - \mathbf{m}_1] = \varepsilon \mathbf{Y}_{\mathbf{m}_1}(\mathbf{m}_2).$$

This symmetry, however, is different for each $\mathbf{m}_1$ so that we may multiplex two such vectors $\mathbf{Y}_{\mathbf{m}_1'}$ and $\mathbf{Y}_{\mathbf{m}_1''}$ into a general *real* vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}_1'} + \mathbf{Y}_{\mathbf{m}_1''},$$

for each of the 2^{n-1} pairs $(\mathbf{m}_1', \mathbf{m}_1'')$. The 2^{n-1} Hermitian-symmetric transform vectors

$$\mathbf{Y}^* = \mathbf{Y}^*_{\mathbf{m}_1'} + \mathbf{Y}^*_{\mathbf{m}_1''}$$

can then be evaluated by the methods of Section 1.3.4.3.5.1(*b*) at the cost of only 2^{n-2} general complex $F(\mathbf{M})$.

The demultiplexing relations by which the separate vectors $\mathbf{Y}^*_{\mathbf{m}_1'}$ and $\mathbf{Y}^*_{\mathbf{m}_1''}$ may be recovered are most simply obtained by observing that the vectors $\mathbf{Z}$ after the twiddle-factor stage are real-valued since $F(2\mathbf{I})$ has a real matrix. Thus, as in Section 1.3.4.3.5.1(*c*)(i),

$$\begin{aligned} \mathbf{Y}^*_{\mathbf{m}_1'} &= (c' - is')\mathbf{R}' \\ \mathbf{Y}^*_{\mathbf{m}_1''} &= (c'' - is'')\mathbf{R}'', \end{aligned}$$

where $\mathbf{R}'$ and $\mathbf{R}''$ are real vectors and where the multipliers $(c' - is')$ and $(c'' - is'')$ are the inverse twiddle factors. Therefore,

$$\begin{aligned} \mathbf{Y}^* &= (c' - is')\mathbf{R}' + (c'' - is'')\mathbf{R}'' \\ &= (c'\mathbf{R}' + c''\mathbf{R}'') - i(s'\mathbf{R}' + s''\mathbf{R}'') \end{aligned}$$

and hence the demultiplexing relation for each $\mathbf{h}_2$:

$$\begin{pmatrix} R' \\ R'' \end{pmatrix} = \frac{1}{c's'' - s'c''} \begin{pmatrix} s'' & -c'' \\ -s' & c' \end{pmatrix} \begin{pmatrix} \mathscr{R}e\, Y^* \\ -\mathscr{I}m\, Y^* \end{pmatrix}.$$

The values of $R'_{\mathbf{h}_2}$ and $R''_{\mathbf{h}_2}$ at those points $\mathbf{h}_2$ where $c's'' - s'c'' = 0$ can be evaluated directly while forming $\mathbf{Y}$. This demultiplexing and the final stage of the calculation, namely

$$F(\mathbf{h}_2 + \mathbf{M}\mathbf{h}_1) = \frac{1}{2^n} \sum_{\mathbf{m}_1 \in \mathbf{Z}^n/2\mathbf{Z}^n} (-1)^{\mathbf{h}_1 \cdot \mathbf{m}_1} R_{\mathbf{m}_1}(\mathbf{h}_2)$$

need only be carried out for the unique half of the range of $\mathbf{h}_2$.

(ii) *Decimation in frequency* ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

Similarly, the vectors $\mathbf{Z}^*_{\mathbf{h}_2}$ of decimated and scrambled results are real and obey internal symmetries

$$\tau_{\mathbf{h}_2} \mathbf{Z}^*_{\mathbf{h}_2} = \varepsilon \check{\mathbf{Z}}^*_{\mathbf{h}_2}$$

which are different for each $\mathbf{h}_2$. For each of the 2^{n-1} pairs $(\mathbf{h}_2', \mathbf{h}_2'')$ the multiplexed vector

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}_2'} + \mathbf{Z}_{\mathbf{h}_2''}$$

is a Hermitian-symmetric vector without internal symmetry, and the 2^{n-1} real vectors

$$\mathbf{Z}^* = \mathbf{Z}^*_{\mathbf{h}_2'} + \mathbf{Z}^*_{\mathbf{h}_2''}$$

may be evaluated at the cost of only 2^{n-2} general complex $F(\mathbf{M})$ by the methods of Section 1.3.4.3.5.1(*c*). The individual transforms $\mathbf{Z}_{\mathbf{h}_2'}$ and $\mathbf{Z}_{\mathbf{h}_2''}$ may then be retrieved *via* the demultiplexing relations

$$\begin{aligned} Z^*_{\mathbf{h}_2'}(\mathbf{h}_1) \quad + Z^*_{\mathbf{h}_2''}(\mathbf{h}_1) \qquad &= Z^*(\mathbf{h}_1) \\ Z^*_{\mathbf{h}_2'}(\mathbf{h}_1 - \mathbf{h}_2') + Z^*_{\mathbf{h}_2''}(\mathbf{h}_1 - \mathbf{h}_2'') &= Z^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1] \end{aligned}$$

which can be solved recursively as described in Section 1.3.4.3.5.1(*b*)(ii). This yields the unique half of the real symmetric results $\mathbf{F}$.

1.3.4.3.5.5. Real antisymmetric transforms

If $\mathbf{X}$ is real antisymmetric, then its transform $\mathbf{X}^*$ is purely imaginary and antisymmetric. The double-multiplexing techniques used for real symmetric transforms may therefore be adapted with only minor changes involving signs and factors of *i*.

1.3.4.3.5.6. Generalized multiplexing

So far the multiplexing technique has been applied to pairs of vectors with similar types of parity-related and/or conjugate symmetry properties, in particular the same value of ε.

It can be generalized so as to accommodate mixtures of vectors with different symmetry characteristics. For example if $\mathbf{X}_1$ is Hermitian-symmetric and $\mathbf{X}_2$ is Hermitian-antisymmetric, so that $\mathbf{X}_1^*$ is real-valued while $\mathbf{X}_2^*$ has purely imaginary values, the multiplexing process should obviously form $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ (instead of $\mathbf{X} = \mathbf{X}_1 + i\mathbf{X}_2$ if both had the same type of symmetry), and demultiplexing consists in separating

$$\begin{aligned} \mathbf{X}_1^* &= \mathscr{R}e\, \mathbf{X}^* \\ \mathbf{X}_2^* &= i\mathscr{I}m\, \mathbf{X}^*. \end{aligned}$$

The general multiplexing formula for pairs of vectors may therefore be written

$$\mathbf{X} = \mathbf{X}_1 + \omega \mathbf{X}_2,$$

where ω is a phase factor (*e.g.* 1 or i) chosen in such a way that all nonexceptional components of $\mathbf{X}_1$ and $\mathbf{X}_2$ (or $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$) be embedded in the complex plane $\mathbb{C}$ along linearly independent directions, thus making multiplexing possible.

It is possible to develop a more general form of multiplexing/demultiplexing for more than two vectors, which can be used to deal with symmetry elements of order 3, 4 or 6. It is based on the theory of group characters (Ledermann, 1987).

1.3.4.3.6. Global crystallographic algorithms

All the necessary ingredients are now available for calculating the CDFT for any given space group.

1.3.4.3.6.1. Triclinic groups

Space group $P1$ is dealt with by the methods of Section 1.3.4.3.5.1 and $P\bar{1}$ by those of Section 1.3.4.3.5.4.

1.3.4.3.6.2. Monoclinic groups

A general monoclinic transformation is of the form

$$S_g : \mathbf{x} \longmapsto \mathbf{R}_g \mathbf{x} + \mathbf{t}_g$$

with $\mathbf{R}_g$ a diagonal matrix whose entries are $+1$ or -1, and $\mathbf{t}_g$ a vector whose entries are 0 or $\frac{1}{2}$. We may thus decompose both real and reciprocal space into a direct sum of a subspace $\mathbb{Z}^{n+}$ where $\mathbf{R}_g$ acts as the identity, and a subspace $\mathbb{Z}^{n-}$ where $\mathbf{R}_g$ acts as minus the identity, with $n_+ + n_- = n = 3$. All usual entities may be correspondingly written as direct sums, for instance:

$$\begin{array}{lll}
\mathbf{R}_g = \mathbf{R}_g^+ \oplus \mathbf{R}_g^-, & \mathbf{N} = \mathbf{N}^+ \oplus \mathbf{N}^-, & \mathbf{M} = \mathbf{M}^+ \oplus \mathbf{M}^-, \\
\mathbf{t}_g = \boldsymbol{t}_g^+ \oplus \mathbf{t}_g^-, & \mathbf{t}_g^{(1)} = \mathbf{t}_g^{(1)+} \oplus \mathbf{t}_g^{(1)-}, & \mathbf{t}_g^{(2)} = \mathbf{t}_g^{(2)+} \oplus \mathbf{t}_g^{(2)-}, \\
\mathbf{m} = \mathbf{m}^+ \oplus \mathbf{m}^-, & \mathbf{m}_1 = \mathbf{m}_1^+ \oplus \mathbf{m}_1^-, & \mathbf{m}_2 = \mathbf{m}_2^+ \oplus \mathbf{m}_2^-, \\
\mathbf{h} = \mathbf{h}^+ \oplus \mathbf{h}^-, & \mathbf{h}_1 = \mathbf{h}_1^+ \oplus \mathbf{h}_1^-, & \mathbf{h}_2 = \mathbf{h}_2^+ \oplus \mathbf{h}_2^-.
\end{array}$$

We will use factoring by 2, with decimation in frequency when computing structure factors, and decimation in time when computing electron densities; this corresponds to $\mathbf{N} = \mathbf{N}_1 \mathbf{N}_2$ with $\mathbf{N}_1 = \mathbf{M}$, $\mathbf{N}_2 = 2\mathbf{I}$. The nonprimitive translation vector $\mathbf{N}\mathbf{t}_g$ then belongs to $\mathbf{M}\mathbb{Z}^n$, and thus

$$\mathbf{t}_g^{(1)} = \mathbf{0} \bmod \mathbf{M}\mathbb{Z}^n, \quad \mathbf{t}_g^{(2)} \in \mathbb{Z}^n / 2\mathbb{Z}^n.$$

The symmetry relations obeyed by ρ and F are as follows: for electron densities

$$\rho(\mathbf{m}^+, \mathbf{m}^-) = \rho(\mathbf{m}^+ + \mathbf{N}^+ \mathbf{t}_g^+, -\mathbf{m}^- - \mathbf{N}^- \mathbf{t}_g^-)$$

or, after factoring by 2,

$$\begin{aligned}
&\rho(\mathbf{m}_1^+, \mathbf{m}_2^+, \mathbf{m}_1^-, \mathbf{m}_2^-) \\
&\quad = \rho(\mathbf{m}_1^+, \mathbf{m}_2^+ + \mathbf{t}_g^{(2)+}, \mathbf{M}^- \boldsymbol{\zeta}(\mathbf{m}_1^-) - \mathbf{m}_1^- - \mathbf{m}_2^-, \mathbf{m}_2^- + \mathbf{t}_g^{(2)-});
\end{aligned}$$

while for structure factors

$$F(\mathbf{h}^+, \mathbf{h}^-) = \exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)] F(\mathbf{h}^+, -\mathbf{h}^-)$$

with its Friedel counterpart

$$F(\mathbf{h}^+, \mathbf{h}^-) = \exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)] \overline{F(-\mathbf{h}^+, \mathbf{h}^-)}$$

or, after factoring by 2,

$$\begin{aligned}
F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-) &= (-1)^{\mathbf{h}_2^+ \cdot \mathbf{t}_g^{(2)+} + \mathbf{h}_2^- \cdot \mathbf{t}_g^{(2)-}} \\
&\quad \times F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{M}^- \boldsymbol{\zeta}(\mathbf{h}_1^-) - \mathbf{h}_1^- - \mathbf{h}_2^-, \mathbf{h}_2^-)
\end{aligned}$$

with Friedel counterpart

$$\begin{aligned}
&F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-) \\
&\quad = (-1)^{\mathbf{h}_2^+ \cdot \mathbf{t}_g^{(2)+} + \mathbf{h}_2^- \cdot \mathbf{t}_g^{(2)-}} \overline{F[\mathbf{M}^+ \boldsymbol{\zeta}(\mathbf{h}_1^+) - \mathbf{h}_1^+ - \mathbf{h}_2^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-]}.
\end{aligned}$$

When calculating electron densities, two methods may be used.

(i) Transform on $\mathbf{h}^-$ first.

The partial vectors defined by $X_{\mathbf{h}^+, \mathbf{h}_2^-} = F(\mathbf{h}^+, \mathbf{h}_1^-, \mathbf{h}_2^-)$ obey symmetry relations of the form

$$X(\mathbf{h}_1^- - \mathbf{h}_2^-) = \varepsilon X[\mathbf{M}^- \boldsymbol{\zeta}(\mathbf{h}_1^-) - \mathbf{h}_1^-]$$

with $\varepsilon = \pm 1$ independent of $\mathbf{h}_1^-$. This is the same relation as for the same parity class of data for a (complex or real) symmetric ($\varepsilon = +1$) or antisymmetric ($\varepsilon = -1$) transform. The same techniques can be used to decrease the number of $F(\mathbf{M}^-)$ by multiplexing pairs of such vectors and demultiplexing their transforms. Partial vectors with different values of ε may be mixed in this way (Section 1.3.4.3.5.6).

Once $F(\mathbf{N}^-)$ is completed, its results have Hermitian symmetry with respect to $\mathbf{h}^+$, and the methods of Section 1.3.4.3.5.1 may be used to obtain the unique electron densities.

(ii) Transform on $\mathbf{h}^+$ first.

The partial vectors defined by $X_{\mathbf{h}^-, \mathbf{h}_2^+} = F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}^-)$ obey symmetry relations of the form

$$X(\mathbf{h}_1^+ - \mathbf{h}_2^+) = \overline{\varepsilon X[\mathbf{M}^+ \boldsymbol{\zeta}(\mathbf{h}_1^+) - \mathbf{h}_1^+]}$$

with $\varepsilon = \pm 1$ independent of $\mathbf{h}_1^+$. This is the same relation as for the same parity class of data for a Hermitian symmetric ($\varepsilon = +1$) or antisymmetric ($\varepsilon = -1$) transform. The same techniques may be used to decrease the number of $F(\mathbf{M}^+)$. This generalizes the procedure described by Ten Eyck (1973) for treating dyad axes, *i.e.* for the case $n_+ = 1$, $\mathbf{t}_g^{(2)-} = \mathbf{0}$, and $\mathbf{t}_g^{(2)+} = \mathbf{0}$ (simple dyad) or $\mathbf{t}_g^{(2)+} \neq \mathbf{0}$ (screw dyad).

Once $F(\mathbf{N}^+)$ is completed, its results have Hermitian symmetry properties with respect to $\mathbf{h}^-$ which can be used to obtain the unique electron densities.

Structure factors may be computed by applying the reverse procedures in the reverse order.

1.3.4.3.6.3. Orthorhombic groups

Almost all orthorhombic space groups are generated by two monoclinic transformations g_1 and g_2 of the type described in Section 1.3.4.3.6.2, with the addition of a centre of inversion $-e$ for centrosymmetric groups. The only exceptions are *Fdd*2 and *Fddd* which contain diamond glides, in which some nonprimitive translations are 'square roots' not of primitive lattice translations, but of centring translations. The generic case will be examined first.

To calculate electron densities, the unique octant of data may first be transformed on $\mathbf{h}^+$ (respectively $\mathbf{h}^-$) as in Section 1.3.4.3.6.2 using the symmetry pertaining to generator g_1. These intermediate results may then be expanded by generator g_2 by the formula of Section 1.3.4.3.3 prior to the final transform on $\mathbf{h}^-$ (respectively $\mathbf{h}^+$). To calculate structure factors, the reverse operations are applied in the reverse order.

The two exceptional groups *Fdd*2 and *Fddd* only require a small modification. The *F*-centring causes the systematic absence of parity classes with mixed parities, leaving only (000) and (111). For the former, the phase factors $\exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)]$ in the symmetry relations of Section 1.3.4.3.6.2 become powers of (-1) so that one is back to the generic case. For the latter, these phase factors are odd powers of i which it is a simple matter to incorporate into a modified multiplexing/demultiplexing procedure.

1.3.4.3.6.4. Trigonal, tetragonal and hexagonal groups

All the symmetries in this class of groups can be handled by the generalized Rader/Winograd algorithms of Section 1.3.4.3.4.3, but no implementation of these is yet available.

In groups containing axes of the form n_m with g.c.d. $(m, n) = 1$ (*i.e.* $3_1, 3_2, 4_1, 4_3, 6_1, 6_5$) along the c direction, the following procedure may be used (Ten Eyck, 1973):

(i) to calculate electron densities, the unique structure factors indexed by

$$[\text{unique } (h, k)] \times (\text{all } l)$$

are transformed on l; the results are rearranged by the transposition formula of Section 1.3.4.3.3 so as to be indexed by

$$[\text{all } (h, k)] \times \left[\text{unique } \left(\frac{1}{n}\right)\text{th of } z\right]$$

and are finally transformed on (h, k) to produce an asymmetric unit. For a dihedral group, the extra twofold axis may be used in the transposition to produce a unique $(1/2n)$th of z.

(ii) to calculate structure factors, the unique densities in $(1/n)$th of z [or $(1/2n)$th for a dihedral group] are first transformed on x and y, then transposed by the formula of Section 1.3.4.3.3 to reindex the intermediate results by

$$[\text{unique } (h, k)] \times (\text{all } z);$$

the last transform on z is then carried out.

1.3.4.3.6.5. Cubic groups

These are usually treated as their orthorhombic or tetragonal subgroups, as the body-diagonal threefold axis cannot be handled by ordinary methods of decomposition.

The three-dimensional factorization technique of Section 1.3.4.3.4.1 allows a complete treatment of cubic symmetry. Factoring by 2 along all three dimensions gives four types (*i.e.* orbits) of parity classes:

(000)	with residual threefold symmetry,
(100), (010), (001)	related by threefold axis,
(110), (101), (011)	related by threefold axis,
(111)	with residual threefold symmetry.

Orbit exchange using the threefold axis thus allows one to reduce the number of partial transforms from 8 to 4 (one per orbit). Factoring by 3 leads to a reduction from 27 to 11 (in this case, further reduction to 9 can be gained by multiplexing the three diagonal classes with residual threefold symmetry into a single class; see Section 1.3.4.3.5.6). More generally, factoring by q leads to a reduction from q^3 to $\frac{1}{3}(q^3 - q) - q$. Each of the remaining transforms then has a symmetry induced from the orthorhombic or tetragonal subgroup, which can be treated as above.

No implementation of this procedure is yet available.

1.3.4.3.6.6. Treatment of centred lattices

Lattice centring is an instance of the duality between periodization and decimation: the extra translational periodicity of ρ induces a decimation of $\mathbf{F} = \{F_\mathbf{h}\}$ described by the 'reflection conditions' on $\mathbf{h}$. As was pointed out in Section 1.3.4.2.2.3, nonprimitive lattices are introduced in order to retain the same matrix representation for a given geometric symmetry operation in all the arithmetic classes in which it occurs. From the computational point of view, therefore, the main advantage in using centred lattices is that it *maximizes decomposability* (Section 1.3.4.2.2.4); reindexing to a primitive lattice would for instance often destroy the diagonal character of the matrix representing a dyad.

In the usual procedure involving three successive one-dimensional transforms, the loss of efficiency caused by the duplication of densities or the systematic vanishing of certain classes of structure factors may be avoided by using a multiplexing/demultiplexing technique (Ten Eyck, 1973):

(i) for base-centred or body-centred lattices, two successive planes of structure factors may be overlaid into a single plane; after transformation, the results belonging to each plane may be separated by parity considerations;

(ii) for face-centred lattices the same method applies, using four successive planes (the third and fourth with an origin translation);

(iii) for rhombohedral lattices in hexagonal coordinates, three successive planes may be overlaid, and the results may be separated by linear combinations involving cube roots of unity.

The three-dimensional factorization technique of Section 1.3.4.3.4.1 is particularly well suited to the treatment of centred lattices: if the decimation matrix of $\mathbf{N}$ contains as a factor $\mathbf{N}_1$ a matrix which 'integerizes' all the nonprimitive lattice vectors, then centring is reflected by the systematic vanishing of certain classes of vectors of decimated data or results, which can simply be omitted from the calculation. An alternative possibly is to reindex on a primitive lattice and use different representative matrices for the symmetry operations: the loss of decomposability is of little consequence in this three-dimensional scheme, although it substantially complicates the definition of the cocycles $\boldsymbol{\mu}_2$ and $\boldsymbol{\eta}_1$.

1.3.4.3.6.7. Programming considerations

The preceding sections have been devoted to showing how the raw computational efficiency of a crystallographic Fourier transform *algorithm* can be maximized. This section will briefly discuss another characteristic (besides speed) which a crystallographic Fourier transform *program* may be required to possess if it is to be useful in various applications: a convenient and versatile mode of presentation of input data or output results.

The standard crystallographic FFT programs (Ten Eyck, 1973, 1985) are rather rigid in this respect, and use rather rudimentary data structures (lists of structure-factor values, and two-dimensional arrays containing successive sections of electron-density maps). It is frequently the case that considerable reformatting of these data or results must be carried out before they can be used in other computations; for instance, maps have to be converted from 2D sections to 3D 'bricks' before they can be inspected on a computer graphics display.

The explicitly three-dimensional approach to the factorization of the DFT and the use of symmetry offers the possibility of richer and more versatile data structures. For instance, the use of 'decimation in frequency' in real space and of 'decimation in time' in reciprocal space leads to data structures in which real-space coordinates are handled by blocks (thus preserving, at least locally, the three-dimensional topological connectivity of the maps) while reciprocal-space indices are handled by parity classes or their generalizations for factors other than 2 (thus making the treatment of centred lattices extremely easy). This global three-dimensional indexing also makes it possible to carry symmetry and multiplicity characteristics for each subvector of intermediate results for the purpose of automating the use of the orbit exchange mechanism.

Brünger (1989) has described the use of a similar three-dimensional factoring technique in the context of structure-factor calculations for the refinement of macromolecular structures.

1.3.4.4. Basic crystallographic computations

1.3.4.4.1. Introduction

Fourier transform (FT) calculations play an indispensable role in crystallography, because the Fourier transformation is inherent in the diffraction phenomenon itself.

Besides this obligatory use, the FT has numerous other applications, motivated more often by its mathematical properties than by direct physical reasoning (although the latter can be supplied after the fact). Typically, many crystallographic computations turn out to be convolutions in disguise, which can be speeded up by orders of magnitude through a judicious use of the FT. Several recent advances in crystallographic computation have been based on this kind of observation.

1.3.4.4.2. Fourier synthesis of electron-density maps

Bragg (1929) was the first to use this type of calculation to assist structure determination. Progress in computing techniques since that time was reviewed in Section 1.3.4.3.1.

The usefulness of the maps thus obtained can be adversely affected by three main factors:

(i) limited resolution;

(ii) errors in the data;

(iii) computational errors.

Limited resolution causes 'series-termination errors' first investigated by Bragg & West (1930), who used an optical analogy with the numerical aperture of a microscope. James (1948*b*) gave a quantitative description of this phenomenon as a convolution with the 'spherical Dirichlet kernel' (Section 1.3.4.2.1.3), which reflects the truncation of the Fourier spectrum by multiplication with the indicator function of the limiting resolution sphere. Bragg & West (1930) suggested that the resulting ripples might be diminished by applying an artificial temperature factor to the data, which performs a further convolution with a Gaussian point-spread function. When the electron-density map is to be used for model refinement, van Reijen (1942) suggested using Fourier coefficients calculated from the model when no observation is available, as a means of combating series-termination effects.

Errors in the data introduce errors in the electron-density maps, with the same mean-square value by virtue of Parseval's theorem. Special positions accrue larger errors (Cruickshank & Rollett, 1953; Cruickshank, 1965*a*). To minimize the mean-square electron-density error due to large phase uncertainties, Blow & Crick (1959) introduced the 'best Fourier' which uses centroid Fourier coefficients; the associated error level in the electron-density map was evaluated by Blow & Crick (1959) and Dickerson *et al.* (1961*a*,*b*).

Computational errors used to be a serious concern when Beevers–Lipson strips were used, and Cochran (1948*a*) carried out a critical evaluation of the accuracy limitations imposed by strip methods. Nowadays, the FFT algorithm implemented on digital computers with a word size of at least 32 bits gives results accurate to six decimal places or better in most applications (see Gentleman & Sande, 1966).

1.3.4.4.3. Fourier analysis of modified electron-density maps

Various approaches to the phase problem are based on certain modifications of the electron-density map, followed by Fourier analysis of the modified map and extraction of phase information from the resulting Fourier coefficients.

1.3.4.4.3.1. Squaring

Sayre (1952*a*) derived his 'squaring method equation' for structures consisting of equal, resolved and spherically symmetric atoms by observing that squaring such an electron density is equivalent merely to sharpening each atom into its square. Thus

$$F_{\mathbf{h}} = \theta_{\mathbf{h}} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}},$$

where $\theta_{\mathbf{h}} = f(\mathbf{h})/f^{\mathrm{sq}}(\mathbf{h})$ is the ratio between the form factor $f(\mathbf{h})$ common to all the atoms and the form factor $f^{\mathrm{sq}}(\mathbf{h})$ for the squared version of that atom.

Most of the central results of direct methods, such as the tangent formula, are an immediate consequence of Sayre's equation. Phase refinement for a macromolecule by enforcement of the squaring method equation was demonstrated by Sayre (1972, 1974).

1.3.4.4.3.2. Other nonlinear operations

A category of phase improvement procedures known as 'density modification' is based on the pointwise application of various quadratic or cubic 'filters' to electron-density maps after removal of negative regions (Hoppe & Gassmann, 1968; Hoppe *et al.*, 1970; Barrett & Zwick, 1971; Gassmann & Zechmeister, 1972; Collins, 1975; Collins *et al.*, 1976; Gassmann, 1976). These operations are claimed to be equivalent to reciprocal-space phase-refinement techniques such as those based on the tangent formula. Indeed the replacement of

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

by $P[\rho(\mathbf{x})]$, where P is a polynomial

$$P(\rho) = a_0 + a_1\rho + a_2\rho^2 + a_3\rho^3 + \ldots$$

yields

$$P[\rho(\mathbf{x})] = a_0 + \sum_{\mathbf{h}} \Bigg[a_1 F_{\mathbf{h}} + a_2 \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} + a_3 \sum_{\mathbf{k}} \sum_{\mathbf{l}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\mathbf{h}-\mathbf{k}-\mathbf{l}} + \ldots \Bigg] \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

and hence gives rise to the convolution-like families of terms encountered in direct methods. This equivalence, however, has been shown to be rather superficial (Bricogne, 1982) because the 'uncertainty principle' embodied in Heisenberg's inequality (Section 1.3.2.4.4.3) imposes severe limitations on the effectiveness of any procedure which operates *pointwise* in *both* real and reciprocal space.

In applying such methods, sampling considerations must be given close attention. If the spectrum of ρ extends to resolution Δ and if the pointwise nonlinear filter involves a polynomial P of

degree n, then $P(\rho)$ should be sampled at intervals of at most $\Delta/2n$ to accommodate the full bandwidth of its spectrum.

1.3.4.4.3.3. Solvent flattening

Crystals of proteins and nucleic acids contain large amounts of mother liquor, often in excess of 50% of the unit-cell volume, occupying connected channels. The well ordered electron density $\rho_{\rm M}(\mathbf{x})$ corresponding to the macromolecule thus occupies only a periodic subregion $\mathscr{U}$ of the crystal. Thus

$$\rho_{\rm M} = \chi_{\mathscr{U}} \times \rho_{\rm M},$$

implying the convolution identity between structure factors (Main & Woolfson, 1963):

$$F_{\rm M}(\mathbf{h}) = \sum_{\mathbf{k}} \bar{\mathscr{F}}\left[\frac{1}{\mathscr{U}}\chi_{\mathscr{U}}\right](\mathbf{h}-\mathbf{k})F_{\rm M}(\mathbf{k})$$

which is a form of the Shannon interpolation formula (Sections 1.3.2.7.1, 1.3.4.2.1.7; Bricogne, 1974; Colman, 1974).

It is often possible to obtain an approximate 'molecular envelope' $\mathscr{U}$ from a poor electron-density map ρ, either interactively by computer graphics (Bricogne, 1976) or automatically by calculating a moving average of the electron density within a small sphere S. The latter procedure can be implemented in real space (Wang, 1985). However, as it is a convolution of ρ with $\chi_{\rm S}$, it can be speeded up considerably (Leslie, 1987) by computing the moving average $\rho_{\rm mav}$ as

$$\rho_{\rm mav}(\mathbf{x}) = \mathscr{T}[\bar{\mathscr{F}}[\rho] \times \bar{\mathscr{F}}[\chi_{\rm S}]](\mathbf{x}).$$

This remark is identical in substance to Booth's method of computation of 'bounded projections' (Booth, 1945*a*) described in Section 1.3.4.2.1.8, except that the summation is kept three-dimensional.

The iterative use of the estimated envelope $\mathscr{U}$ for the purpose of phase improvement (Wang, 1985) is a submethod of the previously developed method of molecular averaging, which is described below. Sampling rules for the Fourier analysis of envelope-truncated maps will be given there.

1.3.4.4.3.4. Molecular averaging by noncrystallographic symmetries

Macromolecules and macromolecular assemblies frequently crystallize with several identical subunits in the asymmetric metric unit, or in several crystal forms containing the same molecule in different arrangements. Rossmann & Blow (1963) recognized that intensity data collected from such structures are redundant (Sayre, 1952*b*) and that their redundancy could be a source of phase information.

The phase constraints implied by the consistency of geometrically redundant intensities were first derived by Rossmann & Blow (1963), and were generalized by Main & Rossmann (1966). Crowther (1967, 1969) reformulated them as linear eigenvalue equations between structure factors, for which he proposed an iterative matrix solution method. Although useful in practice (Jack, 1973), this reciprocal-space approach required computations of size $\propto N^2$ for N reflections, so that N could not exceed a few thousands.

The theory was then reformulated in real space (Bricogne, 1974), showing that the most costly step in Crowther's procedure could be carried out much more economically by averaging the electron densities of all crystallographically independent subunits, then rebuilding the crystal(s) from this averaged subunit, flattening the density in the solvent region(s) by resetting it to its average value. This operation is a *projection* [by virtue of Section 1.3.4.2.2.2(*d*)]. The overall complexity was thus reduced from N^2 to $N \log N$. The design and implementation of a general-purpose program package for averaging, reconstructing and solvent-flattening electron-density maps (Bricogne, 1976) led rapidly to the first high-resolution determinations of virus structures (Bloomer *et al.*, 1978; Harrison *et al.*, 1978), with $N \sim 200\,000$.

The considerable gain in speed is a consequence of the fact that the masking operations used to retrieve the various copies of the common subunit are carried out by simple *pointwise multiplication* by an indicator function χ_U in real space, whereas they involve a *convolution* with $\bar{\mathscr{F}}[\chi_U]$ in reciprocal space.

The averaging by noncrystallographic symmetries of an electron-density map calculated by FFT – hence sampled on a grid which is an integral subdivision of the period lattice – necessarily entails the interpolation of densities at nonintegral points of that grid. The effect of interpolation on the structure factors recalculated from an averaged map was examined by Bricogne (1976). This study showed that, if linear interpolation is used, the initial map should be calculated on a fine grid, of size $\Delta/5$ or $\Delta/6$ at resolution Δ (instead of the previously used value of $\Delta/3$). The analysis about to be given applies to all interpolation schemes which consist in a *convolution* of the sampled density with a fixed *interpolation kernel* function K.

Let ρ be a $\mathbb{Z}^3$-periodic function. Let K be the interpolation kernel in 'normalized' form, *i.e.* such that $\int_{\mathbb{R}^3} K(\mathbf{x})\,{\rm d}^3\mathbf{x} = 1$ and scaled so as to interpolate between sample values given on a unit grid $\mathbb{Z}^3$; in the case of linear interpolation, K is the 'trilinear wedge'

$$K(\mathbf{x}) = W(x)W(y)W(z),$$

where

$$\begin{aligned} W(t) &= 1 - |t| \quad &\text{if } |t| \le 1,\\ &= 0 &\text{if } |t| \ge 1.\end{aligned}$$

Let ρ be sampled on a grid $\mathscr{G}_1 = \mathbf{N}_1^{-1}\mathbb{Z}^3$, and let $I_{\mathbf{N}_1}\rho$ denote the function interpolated from this sampled version of ρ. Then:

$$I_{\mathbf{N}_1}\rho = \left[\rho \times \sum_{\mathbf{m}\in\mathbb{Z}^3} \delta_{(\mathbf{N}_1^{-1}\mathbf{m})}\right] * [(\mathbf{N}_1^{-1})^{\#}K],$$

where $[(N_1^{-1})^{\#}K](\mathbf{x}) = K(\mathbf{N}_1\mathbf{x})$, so that

$$\begin{aligned}\bar{\mathscr{F}}[I_{\mathbf{N}_1}\rho] &= \left[\bar{\mathscr{F}}[\rho] * |\det \mathbf{N}_1| \sum_{\mathbf{k}_1\in\mathbb{Z}^3} \delta_{(\mathbf{N}_1^T\mathbf{k}_1)}\right]\\ &\quad\times \left[\frac{1}{|\det \mathbf{N}_1|}(N_1^T)^{\#}\bar{\mathscr{F}}[K]\right]\\ &= \left[\sum_{\mathbf{k}_1\in\mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_1^T\mathbf{k}_1}\bar{\mathscr{F}}[\rho]\right] \times (N_1^T)^{\#}\bar{\mathscr{F}}[K].\end{aligned}$$

The transform of $I_{\mathbf{N}_1}\rho$ thus consists of

(i) a 'main band' corresponding to $\mathbf{k}_1 = \mathbf{0}$, which consists of the true transform $\bar{\mathscr{F}}[\rho](\boldsymbol{\xi})$ attenuated by multiplication by the central region of $\bar{\mathscr{F}}[K][(\mathbf{N}^{-1})^T\boldsymbol{\xi}]$; in the case of linear interpolation, for example,

$$\bar{\mathscr{F}}[K](\xi,\eta,\zeta) = \left(\frac{\sin\pi\xi}{\pi\xi}\right)^2\left(\frac{\sin\pi\eta}{\pi\eta}\right)^2\left(\frac{\sin\pi\zeta}{\pi\zeta}\right)^2;$$

(ii) a series of 'ghost bands' corresponding to $\mathbf{k}_1 \neq \mathbf{0}$, which consist of translates of $\bar{\mathscr{F}}[\rho]$ multiplied by the tail regions of $(N_1^T)^{\#}\bar{\mathscr{F}}[K]$.

Thus $I_{\mathbf{N}_1}\rho$ is not band-limited even if ρ is. Supposing, however, that ρ is band-limited and that grid $\mathscr{G}_1$ satisfies the Shannon sampling criterion, we see that there will be no overlap between the different bands: $\bar{\mathscr{F}}[\rho]$ may therefore be recovered from the main band by compensating its attenuation, which is approximately a temperature-factor correction.

For numerical work, however, $I_{\mathbf{N}_1}\rho$ must be resampled onto another grid $\mathscr{G}_2$, which causes its transform to become periodized into

$$|\det \mathbf{N}_2| \sum_{\mathbf{k}_2 \in \mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_2^T\mathbf{k}_2} \left\{ \left[\sum_{\mathbf{k}_1 \in \mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_1^T\mathbf{k}_1} \bar{\mathscr{F}}[\rho] \right] (N_1^T)^{\#} \bar{\mathscr{F}}[K] \right\}.$$

This now causes the main band $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{0}$ to become contaminated by the ghost bands ($\mathbf{k}_1 \neq \mathbf{0}$) of the translates ($\mathbf{k}_2 \neq \mathbf{0}$) of $I_{\mathbf{N}_1}\rho$.

Aliasing errors may be minimized by increasing the sampling rate in grid $\mathscr{G}_1$ well beyond the Shannon minimum, which rapidly reduces the r.m.s. content of the ghost bands.

The sampling rate in grid $\mathscr{G}_2$ needs only exceed the Shannon minimum to the extent required to accommodate the increase in bandwidth due to convolution with $\bar{\mathscr{F}}[\chi_U]$, which is the reciprocal-space counterpart of envelope truncation (or solvent flattening) in real space.

1.3.4.4.3.5. Molecular-envelope transforms via Green's theorem

Green's theorem stated in terms of distributions (Section 1.3.2.3.9.1) is particularly well suited to the calculation of the Fourier transforms $\bar{\mathscr{F}}[\chi_U]$ of indicator functions. Let f be the indicator function χ_U and let S be the boundary of U (assumed to be a smooth surface). The jump σ_0 in the value of f across S along the outer normal vector is $\sigma_0 = -1$, the jump σ_ν in the normal derivative of f across S is $\sigma_\nu = 0$, and the Laplacian of f as a function is (almost everywhere) 0 so that $T_{\Delta f} = 0$. Green's theorem then reads:

$$\begin{aligned}\Delta(T_f) &= T_{\Delta f} + \sigma_\nu \delta_{(S)} + \partial_\nu[\sigma_0 \delta_{(S)}] \\ &= -\partial_\nu[\delta_{(S)}].\end{aligned}$$

The function $e_{\mathbf{H}}(\mathbf{X}) = \exp(2\pi i \mathbf{H}\cdot\mathbf{X})$ satisfies the identity $\Delta e_{\mathbf{H}} = -4\pi^2\|\mathbf{H}\|^2 e_{\mathbf{H}}$. Therefore, in Cartesian coordinates:

$$\begin{aligned}\bar{F}[\chi_U](\mathbf{H}) &= \langle T_{\chi_U}, e_{\mathbf{H}} \rangle \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \langle T_{\chi_U}, \Delta e_{\mathbf{H}} \rangle \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \langle \Delta(T_{\chi_U}), e_{\mathbf{H}} \rangle \qquad \text{[Section 1.3.2.3.9.1}(a)\text{]} \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \langle -\partial_\nu[\delta_{(S)}], e_{\mathbf{H}} \rangle \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \int_S \partial_\nu e_{\mathbf{H}} \, \mathrm{d}^2 S \qquad \text{[Section 1.3.2.3.9.1}(c)\text{]} \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \int_S 2\pi i \mathbf{H}\cdot\mathbf{n} \exp(2\pi i \mathbf{H}\cdot\mathbf{X}) \, \mathrm{d}^2 S,\end{aligned}$$

i.e.

$$\bar{\mathscr{F}}[\chi_U](\mathbf{H}) = \frac{1}{2\pi i\|\mathbf{H}\|^2} \int_S \mathbf{H}\cdot\mathbf{n} \exp(2\pi i \mathbf{H}\cdot\mathbf{X}) \, \mathrm{d}^2 S,$$

where $\mathbf{n}$ is the outer normal to S. This formula was used by von Laue (1936) for a different purpose, namely to calculate the transforms of crystal shapes (see also Ewald, 1940). If the surface S is given by a triangulation, the surface integral becomes a sum over all faces, since $\mathbf{n}$ is constant on each face. If U is a solid sphere with radius R, an integration by parts gives immediately:

$$\frac{1}{\mathrm{vol}(U)} \bar{\mathscr{F}}[\chi_U](\mathbf{H}) = \frac{3}{X^3}[\sin X - X\cos X] \quad \text{with } X = 2\pi\|\mathbf{H}\|R.$$

1.3.4.4.4. Structure factors from model atomic parameters

An atomic model of a crystal structure consists of a list of symmetry-unique atoms described by their positions, their thermal agitation and their chemical identity (which can be used as a pointer to form-factor tables). Form factors are usually parameterized as sums of Gaussians, and thermal agitation by a Gaussian temperature factor or tensor. The formulae given in Section 1.3.4.2.2.6 for Gaussian atoms are therefore adequate for most purposes. High-resolution electron-density studies use more involved parameterizations.

Early calculations were carried out by means of Bragg–Lipson charts (Bragg & Lipson, 1936) which gave a graphical representation of the symmetrized trigonometric sums Ξ of Section 1.3.4.2.2.9. The approximation of form factors by Gaussians goes back to the work of Vand *et al.* (1957) and Forsyth & Wells (1959). Agarwal (1978) gave simplified expansions suitable for medium-resolution modelling of macromolecular structures.

This method of calculating structure factors is expensive because *each* atom sends contributions of essentially equal magnitude to *all* structure factors in a resolution shell. The calculation is therefore of size $\propto N\mathscr{N}$ for N atoms and $\mathscr{N}$ reflections. Since N and $\mathscr{N}$ are roughly proportional at a given resolution, this method is very costly for large structures.

Two distinct programming strategies are available (Rollett, 1965) according to whether the fast loop is on all atoms for each reflection, or on all reflections for each atom. The former method was favoured in the early times when computers were unreliable. The latter was shown by Burnett & Nordman (1974) to be more amenable to efficient programming, as no multiplication is required in calculating the arguments of the sine/cosine terms: these can be accumulated by integer addition, and used as subscripts in referencing a trigonometric function table.

1.3.4.4.5. Structure factors via model electron-density maps

Robertson (1936*b*) recognized the similarity between the calculation of structure factors by Fourier summation and the calculation of Fourier syntheses, the main difference being of course that atomic coordinates do not usually lie exactly on a grid obtained by integer subdivision of the crystal lattice. He proposed to address this difficulty by the use of his sorting board, which could extend the scale of subdivision and thus avoid phase errors. In this way the calculation of structure factors became amenable to Beevers–Lipson strip methods, with considerable gain of speed.

Later, Beevers & Lipson (1952) proposed that trigonometric functions attached to atomic positions falling between the grid points on which Beevers–Lipson strips were based should be obtained by linear interpolation from the values found on the strips for the closest grid points. This amounts (Section 1.3.4.4.3.4) to using atoms in the shape of a trilinear wedge, whose

form factor was indicated in Section 1.3.4.4.3.4 and gives rise to aliasing effects (see below) not considered by Beevers & Lipson.

The correct formulation of this idea came with the work of Sayre (1951), who showed that structure factors could be calculated by Fourier analysis of a sampled electron-density map previously generated on a subdivision $\mathbf{N}^{-1}\Lambda$ of the crystal lattice Λ. When generating such a map, care must be taken to distribute onto the sample grid not only the electron densities of all the atoms in the asymmetric motif, but also those of their images under space-group symmetries and lattice translations. Considerable savings in computation occur, especially for large structures, because atoms are *localized*: each atom sends contributions to only a few grid points in real space, rather than to all reciprocal-lattice points. The generation of the sampled electron-density map is still of complexity $\propto N\mathcal{N}$ for N atoms and $\mathcal{N}$ reflections, but the proportionality constant is smaller than that in Section 1.3.4.4.4 by orders of magnitude; the extra cost of Fourier analysis, proportional to $\mathcal{N}\log\mathcal{N}$, is negligible.

The idea of approximating a Fourier transform by a discrete transform on sampled values had already been used by Whittaker (1948), who tested it on the first three odd Hermite functions and did not consider the problem of aliasing errors. By contrast, Sayre gave a lucid analysis of the sampling problems associated to this technique. If the periodic sampled map is written in the form of a weighted lattice distribution (as in Section 1.3.2.7.3) as

$$\rho^{\mathrm{s}} = \sum_{\mathbf{m}\in\mathbb{Z}^3} \rho(\mathbf{N}^{-1}\mathbf{m})\delta_{(\mathbf{N}^{-1}\mathbf{m})},$$

then its discrete Fourier transform yields

$$F^{\mathrm{s}}(\mathbf{h}) = \sum_{\boldsymbol{\eta}\in\mathbb{Z}^3} F(\mathbf{h}+\mathbf{N}^T\boldsymbol{\eta})$$

so that each correct value $F(\mathbf{h})$ is corrupted by its aliases $F(\mathbf{h}+\mathbf{N}^T\boldsymbol{\eta})$ for $\boldsymbol{\eta}\neq\mathbf{0}$.

To cure this aliasing problem, Sayre used 'hypothetical atoms' with form factors equal to those of standard atoms within the resolution range of interest, but set to zero outside that range. This amounts to using atomic densities with built-in series-termination errors, which has the detrimental effect of introducing slowly decaying ripples around the atom which require incrementing sample densities at many more grid points per atom.

Sayre considered another cure in the form of an artificial temperature factor B (Bragg & West, 1930) applied to all atoms. This spreads each atom on more grid points in real space but speeds up the decay of its transform in reciprocal space, thus allowing the use of a coarser sampling grid in real space. He discounted it as spoiling the agreement with observed data, but Ten Eyck (1977) pointed out that this agreement could be restored by applying the negative of the artificial temperature factor to the results. This idea cannot be carried to extremes: if B is chosen too large, the atoms will be so spread out in real space as each to occupy a sizeable fraction of the unit cell and the advantage of atom localization will be lost; furthermore, the form factors will fall off so rapidly that round-off error amplification will occur when the results are sharpened back. Clearly, there exists an optimal combination of B and sampling rate yielding the most economical computation for a given accuracy at a given resolution, and a formula will now be given to calculate it.

Let us make the simplifying assumption that all atoms are roughly equal and that their common form factor can be represented by an equivalent temperature factor B_{eq}. Let $\Delta = 1/d^*_{\max}$ be the resolution to which structure factors are wanted. The Shannon sampling interval is $\Delta/2 = 1/2d^*_{\max}$. Let σ be the oversampling rate, so that the actual sampling interval in the map is $\Delta/2\sigma = 1/2\sigma d^*_{\max}$: then consecutive copies of the transform are separated by a distance $2\sigma d^*_{\max}$ in reciprocal space. Let the artificial temperature factor B_{extra} be added, and let

$$B = B_{\mathrm{eq}} + B_{\mathrm{extra}}.$$

The worst aliasing occurs at the outer resolution limit $d^*_{\max}$, where the 'signal' due to an atom is proportional to

$$\exp[(-B/4)(d^*_{\max})^2],$$

while the 'noise' due to the closest alias is proportional to

$$\exp\{(-B/4)[(2\sigma-1)d^*_{\max}]^2\}.$$

Thus the signal-to-noise ratio, or quality factor, Q is

$$\exp[B\sigma(\sigma-1)(d^*_{\max})^2].$$

If a certain value of Q is desired (*e.g.* $Q = 100$ for 1% accuracy), then the equation

$$B = \frac{\log Q}{\sigma(\sigma-1)(d^*_{\max})^2}$$

defines B in terms of σ, $d^*_{\max}$ and Q.

The overall cost of the structure-factor calculation from N atoms is then

(i) $C_1 \times B^{2/3} \times N$ for density generation,

(ii) $C_2 \times (2\sigma d^*_{\max})^3 \times \log[(2\sigma d^*_{\max})^3]$ for Fourier analysis,

where C_1 and C_2 are constant depending on the speed of the computer used. This overall cost may be minimized with respect to σ for given $d^*_{\max}$ and Q, determining the optimal B (and hence B_{extra}) in passing by the above relation.

Sayre (1951) did observe that applying an artificial temperature factor in real space would not create series-termination ripples: the resulting atoms would have a smaller effective radius than his hypothetical atoms, so that step (i) would be faster. This optimality of Gaussian smearing is ultimately a consequence of Hardy's theorem (Section 1.3.2.4.4.3).

1.3.4.4.6. Derivatives for variational phasing techniques

Some methods of phase determination rely on maximizing a certain global criterion $S[\rho]$ involving the electron density, of the form $\int_{\mathbb{R}^3/\mathbb{Z}^3} K[\rho(\mathbf{x})]\,\mathrm{d}^3\mathbf{x}$, under constraint of agreement with the observed structure-factor amplitudes, typically measured by a χ^2 residual C. Several recently proposed methods use for $S[\rho]$ various measures of entropy defined by taking $K(\rho) = -\rho\log(\rho/\mu)$ or $K(\rho) = \log\rho$ (Bricogne, 1982; Britten & Collins, 1982; Narayan & Nityananda, 1982; Bryan *et al.*, 1983; Wilkins *et al.*, 1983; Bricogne, 1984; Navaza, 1985; Livesey & Skilling, 1985). Sayre's use of the squaring method to improve protein phases (Sayre, 1974) also belongs to this category, and is amenable to the same computational strategies (Sayre, 1980).

These methods differ from the density-modification procedures of Section 1.3.4.4.3.2 in that they seek an optimal solution by moving electron densities (or structure factors) *jointly* rather than pointwise, *i.e.* by moving along suitably chosen *search directions* $v_i(\mathbf{x})$ [or $V_i(\mathbf{h})$].

For computational purposes, these search directions may be handled either as column vectors of sample values $\{v_i(\mathbf{N}^{-1}\mathbf{m})\}_{\mathbf{m}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3}$ on a grid in real space, or as column vectors of Fourier coefficients $\{V_i(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3}$ in reciprocal space. These column vectors are the coordinates of the same vector $\mathbf{V}_i$ in an

abstract vector space $\mathscr{V} \cong L(\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3)$ of dimension $\mathscr{N} = |\det \mathbf{N}|$ over $\mathbb{R}$, but referred to two different bases which are related by the DFT and its inverse (Section 1.3.2.7.3).

The problem of finding the optimum of S for a given value of C amounts to achieving collinearity between the gradients ∇S and ∇C of S and of C in $\mathscr{V}$, the scalar ratio between them being a Lagrange multiplier. In order to move towards such a solution from a trial position, the dependence of ∇S and ∇C on position in $\mathscr{V}$ must be represented. This involves the $\mathscr{N} \times \mathscr{N}$ Hessian matrices $\mathbf{H}(S)$ and $\mathbf{H}(C)$, whose size precludes their use in the whole of $\mathscr{V}$. Restricting the search to a smaller *search subspace* of dimension n spanned by $\{\mathbf{V}_i\}_{i=1,\ldots,n}$ we may build local quadratic models of S and C (Bryan & Skilling, 1980; Burch *et al.*, 1983) with respect to n coordinates $\mathbf{X}$ in that subspace:

$$\begin{aligned} S(\mathbf{X}) &= S(\mathbf{X}_0) + \mathbf{S}_0^T(\mathbf{X} - \mathbf{X}_0) \\ &\quad + \tfrac{1}{2}(\mathbf{X} - \mathbf{X}_0)^T \mathbf{H}_0(S)(\mathbf{X} - \mathbf{X}_0) \\ C(\mathbf{X}) &= C(\mathbf{X}_0) + \mathbf{C}_0^T(\mathbf{X} - \mathbf{X}_0) \\ &\quad + \tfrac{1}{2}(\mathbf{X} - \mathbf{X}_0)^T \mathbf{H}_0(C)(\mathbf{X} - \mathbf{X}_0). \end{aligned}$$

The coefficients of these linear models are given by scalar products:

$$\begin{aligned} [\mathbf{S}_0]_i &= (\mathbf{V}_i, \nabla S) \\ [\mathbf{C}_0]_i &= (\mathbf{V}_i, \nabla C) \\ [\mathbf{H}_0(S)]_{ij} &= [\mathbf{V}_i, \mathbf{H}(S)\mathbf{V}_j] \\ [\mathbf{H}_0(C)]_{ij} &= [\mathbf{V}_i, \mathbf{H}(C)\mathbf{V}_j] \end{aligned}$$

which, by virtue of Parseval's theorem, may be evaluated either in real space or in reciprocal space (Bricogne, 1984). In doing so, special positions and reflections must be taken into account, as in Section 1.3.4.2.2.8. Scalar products involving S are best evaluated by real-space grid summation, because $\mathbf{H}(S)$ is *diagonal* in this representation; those involving C are best calculated by reciprocal-space summation, because $\mathbf{H}(C)$ is at worst 2×2 *block-diagonal* in this representation. Using these Hessian matrices in the wrong space would lead to prohibitively expensive convolutions instead of scalar (or at worst 2×2 matrix) multiplications.

1.3.4.4.7. Derivatives for model refinement

Since the origins of X-ray crystal structure analysis, the calculation of crystallographic Fourier series has been closely associated with the process of refinement. Fourier coefficients *with phases* were obtained for all or part of the measured reflections in the basis of some trial model for all or part of the structure, and Fourier syntheses were then used to complete and improve this initial model. This approach is clearly described in the classic paper by Bragg & West (1929), and was put into practice in the determination of the structures of topaz (Alston & West, 1929) and diopside (Warren & Bragg, 1929). Later, more systematic methods of arriving at a trial model were provided by the Patterson synthesis (Patterson, 1934, 1935*a*,*b*; Harker, 1936) and by isomorphous replacement (Robertson, 1935, 1936*c*). The role of Fourier syntheses, however, remained essentially unchanged [see Robertson (1937) for a review] until more systematic methods of structure refinement were introduced in the 1940s. A particularly good account of the processes of structure completion and refinement may be found in Chapters 15 and 16 of Stout & Jensen (1968).

It is beyond the scope of this section to review the vast topic of refinement methods: rather, it will give an account of those aspects of their development which have sought improved power by exploiting properties of the Fourier transformation. It is of more than historical interest that some recent advances in the crystallographic refinement of macromolecular structures had been anticipated by Cochran and Cruickshank in the early 1950s.

1.3.4.4.7.1. The method of least squares

Hughes (1941) was the first to use the already well established multivariate least-squares method (Whittaker & Robinson, 1944) to refine initial estimates of the parameters describing a model structure. The method gained general acceptance through the programming efforts of Friedlander *et al.* (1955), Sparks *et al.* (1956), Busing & Levy (1961) and others.

The Fourier relations between ρ and F (Section 1.3.4.2.2.6) are used to derive the 'observational equations' connecting the structure parameters $\{u_p\}_{p=1,\ldots,n}$ to the observations $\{|F_\mathbf{h}|^{\rm obs}, (\sigma_\mathbf{h}^2)^{\rm obs}\}_{\mathbf{h}\in\mathscr{H}}$ comprising the amplitudes and their experimental variances for a set $\mathscr{H}$ of unique reflections.

The normal equations giving the corrections $\delta\mathbf{u}$ to the parameters are then

$$(\mathbf{A}^T\mathbf{W}\mathbf{A})\delta\mathbf{u} = -\mathbf{A}^T\mathbf{W}\Delta,$$

where

$$\begin{aligned} A_{\mathbf{h}p} &= \frac{\partial |F_\mathbf{h}^{\rm calc}|}{\partial u_p} \\ \Delta_\mathbf{h} &= |F_\mathbf{h}^{\rm calc}| - |F_\mathbf{h}|^{\rm obs} \\ \mathbf{W} &= \text{diag } (W_\mathbf{h}) \quad \text{with} \quad W_\mathbf{h} = \frac{1}{(\sigma_\mathbf{h}^2)^{\rm obs}}. \end{aligned}$$

To calculate the elements of $\mathbf{A}$, write:

$$F = |F| \exp(i\varphi) = \alpha + i\beta;$$

hence

$$\begin{aligned} \frac{\partial |F|}{\partial u} &= \frac{\partial \alpha}{\partial u}\cos\varphi + \frac{\partial \beta}{\partial u}\sin\varphi \\ &= \mathscr{R}e\left[\frac{\partial F}{\partial u}\overline{\exp(i\varphi)}\right] = \mathscr{R}e\left[\frac{\overline{\partial F}}{\partial u}\exp(i\varphi)\right]. \end{aligned}$$

In the simple case of atoms with real-valued form factors and isotropic thermal agitation in space group $P1$,

$$F_\mathbf{h}^{\rm calc} = \sum_{j\in J} g_j(\mathbf{h}) \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_j),$$

where

$$g_j(\mathbf{h}) = Z_j f_j(\mathbf{h}) \exp[-\tfrac{1}{4}B_j(d_\mathbf{h}^*)^2],$$

Z_j being a fractional occupancy.

Positional derivatives with respect to $\mathbf{x}_j$ are given by

$$\begin{aligned} \frac{\partial F_\mathbf{h}^{\rm calc}}{\partial \mathbf{x}_j} &= (2\pi i\mathbf{h}) g_j(\mathbf{h}) \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_j) \\ \frac{\partial |F_\mathbf{h}^{\rm calc}|}{\partial \mathbf{x}_j} &= \mathscr{R}e[(-2\pi i\mathbf{h}) g_j(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}_j) \exp(i\varphi_\mathbf{h}^{\rm calc})] \end{aligned}$$

so that the corresponding 3×1 subvector of the right-hand side of the normal equations reads:

$$-\sum_{\mathbf{h}\in\mathscr{H}} W_{\mathbf{h}} \frac{\partial|F_{\mathbf{h}}^{\rm calc}|}{\partial \mathbf{x}_j}(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) = -\mathscr{R}e\Bigg[\sum_{\mathbf{h}\in\mathscr{H}} g_j(\mathbf{h})(-2\pi i\mathbf{h})W_{\mathbf{h}}(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \times \exp(i\varphi_{\mathbf{h}}^{\rm calc})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)\Bigg].$$

The setting up and solution of the normal equations lends itself well to computer programming and has the advantage of providing a thorough analysis of the accuracy of its results (Cruickshank, 1965*b*, 1970; Rollett, 1970). It is, however, an expensive task, of complexity $\propto n \times |\mathscr{H}|^2$, which is unaffordable for macromolecules.

1.3.4.4.7.2. Booth's differential Fourier syntheses

It was the use of Fourier syntheses in the completion of trial structures which provided the incentive to find methods for computing 2D and 3D syntheses efficiently, and led to the Beevers–Lipson strips. The limited accuracy of the latter caused the estimated positions of atoms (identified as peaks in the maps) to be somewhat in error. Methods were therefore sought to improve the accuracy with which the coordinates of the electron-density maxima could be determined. The naive method of peak-shape analysis from densities recalculated on a $3 \times 3 \times 3$ grid using high-accuracy trigonometric tables entailed 27 summations per atom.

Booth (1946*a*) suggested examining the rapidly varying derivatives of the electron density rather than its slowly varying values. If

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i\mathbf{h}\cdot\mathbf{x})$$

then the gradient vector $\nabla_{\mathbf{x}}\rho$ of ρ at $\mathbf{x}^0$

$$(\nabla_{\mathbf{x}}\rho)(\mathbf{x}^0) = \sum_{\mathbf{h}} F_{\mathbf{h}}(-2\pi i\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}^0)$$

can be calculated by means of three Fourier summations from the 3×1 vector of Fourier coefficients

$$(-2\pi i\mathbf{h})F_{\mathbf{h}}.$$

Similarly, the Hessian matrix of ρ at $\mathbf{x}^0$

$$[(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T)\rho](\mathbf{x}^0) = \sum_{\mathbf{h}} F_{\mathbf{h}}(-4\pi^2\mathbf{h}\mathbf{h}^T)\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}^0)$$

can be calculated by six Fourier summations from the unique elements of the symmetric matrix of Fourier coefficients:

$$-4\pi^2\begin{pmatrix} h^2 & hk & hl \\ hk & k^2 & kl \\ hl & kl & l^2 \end{pmatrix}F_{\mathbf{h}}.$$

The scalar maps giving the components of the gradient and Hessian matrix of ρ will be called differential syntheses of 1st order and 2nd order respectively. If $\mathbf{x}^0$ is approximately but not exactly a maximum of ρ, then the Newton–Raphson estimate of the true maximum $\mathbf{x}^*$ is given by:

$$\mathbf{x}^* = \mathbf{x}^0 - [[(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T)\rho](\mathbf{x}^0)]^{-1}[\nabla_{\mathbf{x}}\rho(\mathbf{x}^0)].$$

This calculation requires only nine accurate Fourier summations (instead of 27), and this number is further reduced to four if the peak is assumed to be spherically symmetrical.

The resulting positions are affected by series-termination errors in the differential syntheses. Booth (1945*c*, 1946*c*) proposed a 'back-shift correction' to eliminate them, and extended this treatment to the acentric case (Booth, 1946*b*). He cautioned against the use of an artificial temperature factor to fight series-termination errors (Brill *et al.*, 1939), as this could be shown to introduce coordinate errors by causing overlap between atoms (Booth, 1946*c*, 1947*a*,*b*).

Cruickshank was able to derive estimates for the standard uncertainties of the atomic coordinates obtained in this way (Cox & Cruickshank, 1948; Cruickshank, 1949*a*,*b*) and to show that they agreed with those provided by the least-squares method.

The calculation of differential Fourier syntheses was incorporated into the crystallographic programs of Ahmed & Cruickshank (1953*b*) and of Sparks *et al.* (1956).

1.3.4.4.7.3. Booth's method of steepest descents

Having defined the now universally adopted R factors (Booth, 1945*b*) as criteria of agreement between observed and calculated amplitudes or intensities, Booth proposed that R should be minimized with respect to the set of atomic coordinates $\{\mathbf{x}_j\}_{j\in J}$ by descending along the gradient of R in parameter space (Booth, 1947*c*,*d*). This 'steepest descents' procedure was compared with Patterson methods by Cochran (1948*d*).

When calculating the necessary derivatives, Booth (1948*a*, 1949) used the formulae given above in connection with least squares. This method was implemented by Qurashi (1949) and by Vand (1948, 1951) with parameter-rescaling modifications which made it very close to the least-squares method (Cruickshank, 1950; Qurashi & Vand, 1953; Qurashi, 1953).

1.3.4.4.7.4. Cochran's Fourier method

Cochran (1948*b*,*c*, 1951*a*) undertook to exploit an algebraic similarity between the right-hand side of the normal equations in the least-squares method on the one hand, and the expression for the coefficients used in Booth's differential syntheses on the other hand (see also Booth, 1948*a*). In doing so he initiated a remarkable sequence of formal and computational developments which are still actively pursued today.

Let $\rho_{\rm C}(\mathbf{x})$ be the electron-density map corresponding to the current atomic model, with structure factors $|F_{\mathbf{h}}^{\rm calc}|\exp(i\varphi_{\mathbf{h}}^{\rm calc})$; and let $\rho_{\rm O}(\mathbf{x})$ be the map calculated from observed moduli and calculated phases, *i.e.* with coefficients $\{|F_{\mathbf{h}}|^{\rm obs}\exp(i\varphi_{\mathbf{h}}^{\rm calc})\}_{\mathbf{h}\in\mathscr{H}}$. If there are enough data for $\rho_{\rm C}$ to have a resolved peak at each model atomic position $\mathbf{x}_j$, then

$$(\nabla_{\mathbf{x}}\rho_{\rm C})(\mathbf{x}_j) = \mathbf{0} \quad \text{for each } j \in J;$$

while if the calculated phases $\varphi_{\mathbf{h}}^{\rm calc}$ are good enough, $\rho_{\rm O}$ will also have peaks at each $\mathbf{x}_j$:

$$(\nabla_{\mathbf{x}}\rho_{\rm O})(\mathbf{x}_j) = \mathbf{0} \quad \text{for each } j \in J.$$

It follows that

$$[\nabla_{\mathbf{x}}(\rho_{\rm C}-\rho_{\rm O})](\mathbf{x}_j)=\sum_{\mathbf{h}}(-2\pi i\mathbf{h})[(|F_{\mathbf{h}}^{\rm calc}|-|F_{\mathbf{h}}|^{\rm obs})\exp(i\varphi_{\mathbf{h}}^{\rm calc})]\times\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)=\mathbf{0}\text{ for each }j\in J,$$

where the summation is over all reflections in $\mathscr{H}$ or related to $\mathscr{H}$ by space-group and Friedel symmetry (overlooking multiplicity factors!). This relation is less sensitive to series-termination errors than either of the previous two, since the spectrum of $\rho_{\rm O}$ could have been extrapolated beyond the data in $\mathscr{H}$ by using that of $\rho_{\rm C}$ [as in van Reijen (1942)] without changing its right-hand side.

Cochran then used the identity

$$\frac{\partial F_{\mathbf{h}}^{\rm calc}}{\partial\mathbf{x}_j}=(2\pi i\mathbf{h})g_j(\mathbf{h})\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j)$$

in the form

$$(-2\pi i\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)=\frac{1}{g_j(\mathbf{h})}\overline{\frac{\partial F_{\mathbf{h}}^{\rm calc}}{\partial\mathbf{x}_j}}$$

to rewrite the previous relation as

$$\begin{aligned}&[\nabla_{\mathbf{x}}(\rho_{\rm C}-\rho_{\rm O})](\mathbf{x}_j)\\&=\sum_{\mathbf{h}}\frac{1}{g_j(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}|-|F_{\mathbf{h}}|^{\rm obs})\mathscr{R}e\left[\overline{\frac{\partial F_{\mathbf{h}}^{\rm calc}}{\partial\mathbf{x}_j}}\exp(i\varphi_{\mathbf{h}}^{\rm calc})\right]\\&=\sum_{\mathbf{h}}\frac{1}{g_j(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}|-|F_{\mathbf{h}}|^{\rm obs})\frac{\partial|F_{\mathbf{h}}^{\rm calc}|}{\partial\mathbf{x}_j}\\&=\mathbf{0}\quad\text{for each }j\in J\end{aligned}$$

(the operation $\mathscr{R}e[]$ on the first line being neutral because of Friedel symmetry). This is equivalent to the vanishing of the 3×1 subvector of the right-hand side of the normal equations associated to a least-squares refinement in which the weights would be

$$W_{\mathbf{h}}=\frac{1}{g_j(\mathbf{h})}.$$

Cochran concluded that, for equal-atom structures with $g_j(\mathbf{h})=g(\mathbf{h})$ for all j, the positions $\mathbf{x}_j$ obtained by Booth's method applied to the difference map $\rho_{\rm O}-\rho_{\rm C}$ are such that they minimize the residual

$$\tfrac{1}{2}\sum_{\mathbf{h}}\frac{1}{g(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}|-|F_{\mathbf{h}}|^{\rm obs})^2$$

with respect to the atomic positions. If it is desired to minimize the residual of the ordinary least-squares method, then the differential synthesis method should be applied to the *weighted* difference map

$$\sum_{\mathbf{h}}\frac{W_{\mathbf{h}}}{g(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}|-|F_{\mathbf{h}}|^{\rm obs})\exp(i\varphi_{\mathbf{h}}^{\rm calc}).$$

He went on to show (Cochran, 1951*b*) that the refinement of temperature factors could also be carried out by inspecting appropriate derivatives of the weighted difference map.

This Fourier method was used by Freer *et al.* (1976) in conjunction with a stereochemical regularization procedure to refine protein structures.

1.3.4.4.7.5. Cruickshank's modified Fourier method

Cruickshank consolidated and extended Cochran's derivations in a series of classic papers (Cruickshank, 1949*b*, 1950, 1952, 1956). He was able to show that all the coefficients involved in the right-hand side and normal matrix of the least-squares method could be calculated by means of suitable differential Fourier syntheses *even when the atoms overlap*. This remarkable achievement lay essentially dormant until its independent rediscovery by Agarwal in 1978 (Section 1.3.4.4.7.6).

To ensure rigorous equivalence between the summations over $\mathbf{h}\in\mathscr{H}$ (in the expressions of least-squares right-hand side and normal matrix elements) and genuine Fourier summations, multiplicity-corrected weights were introduced by:

$$w_{\mathbf{h}}=\frac{1}{|G_{\mathbf{h}}|}W_{\mathbf{h}}\quad\text{if }\mathbf{h}\in G\mathbf{h}\quad\text{with }\mathbf{h}\in\mathscr{H},$$

$$w_{\mathbf{h}}=0\qquad\text{otherwise},$$

where $G\mathbf{h}$ denotes the orbit of $\mathbf{h}$ and $G_{\mathbf{h}}$ its isotropy subgroup (Section 1.3.4.2.2.5). Similarly, derivatives with respect to parameters of symmetry-unique atoms were expressed, *via* the chain rule, as sums over the orbits of these atoms.

Let $p=1,\ldots,n$ be the label of a parameter u_p belonging to atoms with label j. Then Cruickshank showed that the pth element of the right-hand side of the normal equations can be obtained as $D_{p,j}(\mathbf{x}_j)$, where $D_{p,j}$ is a differential synthesis of the form

$$D_{p,j}(\mathbf{x})=\sum_{\mathbf{h}}P_p(\mathbf{h})g_j(\mathbf{h})w_{\mathbf{h}}(|F_{\mathbf{h}}^{\rm calc}|-|F_{\mathbf{h}}|^{\rm obs})\times\exp(i\varphi_{\mathbf{h}}^{\rm calc})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x})$$

with $P_p(\mathbf{h})$ a polynomial in (h, k, l) depending on the type of parameter p. The correspondence between parameter type and the associated polynomial extends Booth's original range of differential syntheses, and is recapitulated in the following table.

Parameter type	$P(h,k,l)$
x coordinate	$-2\pi ih$
y coordinate	$-2\pi ik$
z coordinate	$-2\pi il$
B isotropic	$-\frac{1}{4}(d_{\mathbf{h}}^*)^2$
B^{11} anisotropic	$-h^2$
B^{12} anisotropic	$-hk$
B^{13} anisotropic	$-hl$
B^{22} anisotropic	$-k^2$
B^{23} anisotropic	$-kl$
B^{33} anisotropic	$-l^2$.

Unlike Cochran's original heuristic argument, this result does not depend on the atoms being resolved.

Cruickshank (1952) also considered the elements of the normal matrix, of the form

$$\sum_{\mathbf{h}} w_{\mathbf{h}} \frac{\partial |F_{\mathbf{h}}^{\rm calc}|}{\partial u_p} \frac{\partial |F_{\mathbf{h}}^{\rm calc}|}{\partial u_q}$$

associated with positional parameters. The 3×3 block for parameters $\mathbf{x}_j$ and $\mathbf{x}_k$ may be written

$$\sum_{\mathbf{h}} w_{\mathbf{h}} (\mathbf{h}\mathbf{h}^T) \mathscr{R}e[(-2\pi i) g_j(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) \exp(i\varphi_{\mathbf{h}}^{\rm calc})] \times \mathscr{R}e[(-2\pi i) g_k(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_k) \exp(i\varphi_{\mathbf{h}}^{\rm calc})]$$

which, using the identity

$$\mathscr{R}e(z_1)\mathscr{R}e(z_2) = \tfrac{1}{2}[\mathscr{R}e(z_1 z_2) + \mathscr{R}e(z_1 \overline{z_2})],$$

becomes

$$2\pi^2 \sum_{\mathbf{h}} w_{\mathbf{h}} (\mathbf{h}\mathbf{h}^T) g_j(\mathbf{h}) g_k(\mathbf{h}) \times \{\exp[-2\pi i \mathbf{h} \cdot (\mathbf{x}_j - \mathbf{x}_k)] - \exp(2i\varphi_{\mathbf{h}}^{\rm calc}) \exp[-2\pi i \mathbf{h} \cdot (\mathbf{x}_j + \mathbf{x}_k)]\}$$

(Friedel's symmetry makes $\mathscr{R}e$ redundant on the last line). Cruickshank argued that the first term would give a good approximation to the diagonal blocks of the normal matrix and to those off-diagonal blocks for which $\mathbf{x}_j$ and $\mathbf{x}_k$ are close. On this basis he was able to justify the '*n*-shift rule' of Shoemaker *et al.* (1950). Cruickshank gave this derivation in a general space group, but using a very terse notation which somewhat obscures it. Using the symmetrized trigonometric structure-factor kernel Ξ^- of Section 1.3.4.2.2.9 and its multiplication formula, the above expression is seen to involve the values of a Fourier synthesis at points of the form $\mathbf{x}_j \pm S_g(\mathbf{x}_k)$.

Cruickshank (1956) showed that this analysis could also be applied to the refinement of temperature factors.

These two results made it possible to obtain all coefficients involved in the normal equations by looking up the values of certain differential Fourier syntheses at $\mathbf{x}_j$ or at $\mathbf{x}_j \pm S_g(\mathbf{x}_k)$. At the time this did not confer any superiority over the standard form of the least-squares procedure, because the accurate computation of Fourier syntheses was an expensive operation. The modified Fourier method was used by Truter (1954) and by Ahmed & Cruickshank (1953*a*), and was incorporated into the program system described by Cruickshank *et al.* (1961). A more recent comparison with the least-squares method was made by Dietrich (1972).

There persisted, however, some confusion about the nature of the relationship between Fourier and least-squares methods, caused by the extra factors $g_j(\mathbf{h})$ which make it necessary to compute a differential synthesis for each type of atom. This led Cruickshank to conclude that 'in spite of their remarkable similarities the least-squares and modified-Fourier methods are fundamentally distinct'.

1.3.4.4.7.6. Agarwal's FFT implementation of the Fourier method

Agarwal (1978) rederived and completed Cruickshank's results at a time when the availability of the FFT algorithm made the Fourier method of calculating the coefficients of the normal equations much more economical than the standard method, especially for macromolecules.

As obtained by Cruickshank, the modified Fourier method required a full 3D Fourier synthesis

– for each type of parameter, since this determines [*via* the polynomial $P_p(\mathbf{h})$] the type of differential synthesis to be computed;

– for each type of atom $j \in J$, since the coefficients of the differential synthesis must be multiplied by $g_j(\mathbf{h})$.

Agarwal disposed of the latter dependence by pointing out that the multiplication involved is equivalent to a real-space *convolution* between the differential synthesis and $\sigma_j(\mathbf{x})$, the standard electron density ρ_j for atom type j (Section 1.3.4.2.1.2) smeared by the isotropic thermal agitation of that atom. Since σ_j is *localized*, this convolution involves only a small number of grid points. The requirement of a distinct differential synthesis for each parameter type, however, continued to hold, and created some difficulties at the FFT level because the symmetries of differential syntheses are more complex than ordinary space-group symmetries. Jack & Levitt (1978) sought to avoid the calculation of difference syntheses by using instead finite differences calculated from ordinary Fourier or difference Fourier maps.

In spite of its complication, this return to the Fourier implementation of the least-squares method led to spectacular increases in speed (Isaacs & Agarwal, 1978; Agarwal, 1980; Baker & Dodson, 1980) and quickly gained general acceptance (Dodson, 1981; Isaacs, 1982*a*,*b*, 1984).

1.3.4.4.7.7. Lifchitz's reformulation

Lifchitz [see Agarwal *et al.* (1981), Agarwal (1981)] proposed that the idea of treating certain multipliers in Cruickshank's modified differential Fourier syntheses by means of a convolution in real space should be applied not only to $g_j(\mathbf{h})$, but also to the polynomials $P_p(\mathbf{h})$ which determine the type of differential synthesis being calculated. This leads to convoluting $\partial\sigma_j/\partial u_p$ with the *same* ordinary weighted difference Fourier synthesis, rather than σ_j with the differential synthesis of type p. In this way, a *single* Fourier synthesis, with *ordinary* (scalar) symmetry properties, needs be computed; the parameter type and atom type both intervene through the function $\partial\sigma_j/\partial u_p$ with which it is convoluted. This approach has been used as the basis of an efficient general-purpose least-squares refinement program for macromolecular structures (Tronrud *et al.*, 1987).

This rearrangement amounts to using the fact (Section 1.3.2.3.9.7) that convolution commutes with differentiation. Let

$$D(\mathbf{x}) = \sum_{\mathbf{h}} w_{\mathbf{h}} (|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \exp(i\varphi_{\mathbf{h}}^{\rm calc}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

be the inverse-variance weighted difference map, and let us assume that parameter u_p belongs to atom j. Then the Agarwal form for the pth component of the right-hand side of the normal equations is

$$\left(\frac{\partial D}{\partial u_p} * \sigma_j\right)(x_j),$$

while the Lifchitz form is

$$\left(D * \frac{\partial \sigma_j}{\partial u_p}\right)(\mathbf{x}_j).$$

1.3.4.4.7.8. A simplified derivation

A very simple derivation of the previous results will now be given, which suggests the possibility of many generalizations.

The weighted difference map $D(\mathbf{x})$ has coefficients $D_{\mathbf{h}}$ which are the gradients of the global residual with respect to each $F_{\mathbf{h}}^{\rm calc}$:

$$D_{\mathbf{h}} = \frac{\partial R}{\partial A_{\mathbf{h}}^{\rm calc}} + i\frac{\partial R}{\partial B_{\mathbf{h}}^{\rm calc}}.$$

By the chain rule, a variation of each $F_{\mathbf{h}}^{\rm calc}$ by $\delta F_{\mathbf{h}}^{\rm calc}$ will result in a variation of R by δR with

$$\delta R = \sum_{\mathbf{h}}\left[\frac{\partial R}{\partial A_{\mathbf{h}}^{\rm calc}}\delta A_{\mathbf{h}}^{\rm calc} + \frac{\partial R}{\partial B_{\mathbf{h}}^{\rm calc}}\delta B_{\mathbf{h}}^{\rm calc}\right] = \mathscr{R}e\sum_{\mathbf{h}}[\overline{D_{\mathbf{h}}}\delta F_{\mathbf{h}}^{\rm calc}].$$

The $\mathscr{R}e$ operation is superfluous because of Friedel symmetry, so that δR may be simply written in terms of the Hermitian scalar product in $\ell^2(\mathbb{Z}^3)$:

$$\delta R = (\mathbf{D}, \delta\mathbf{F}^{\rm calc}).$$

If $\rho^{\rm calc}$ is the transform of $\delta\mathbf{F}^{\rm calc}$, we have also by Parseval's theorem

$$\delta R = (D, \delta\rho^{\rm calc}).$$

We may therefore write

$$D(\mathbf{x}) = \frac{\partial R}{\partial\rho^{\rm calc}(\mathbf{x})},$$

which states that $D(\mathbf{x})$ is the functional derivative of R with respect to $\rho^{\rm calc}$.

The right-hand side of the normal equations has $\partial R/\partial u_p$ for its pth element, and this may be written

$$\frac{\partial R}{\partial u_p} = \int_{\mathbb{R}^3/\mathbb{Z}^3}\frac{\partial R}{\partial\rho^{\rm calc}(\mathbf{x})}\frac{\partial\rho^{\rm calc}(\mathbf{x})}{\partial u_p}{\rm d}^2\mathbf{x} = \left(D, \frac{\partial\rho^{\rm calc}}{\partial u_p}\right).$$

If u_p belongs to atom j, then

$$\frac{\partial\rho^{\rm calc}}{\partial u_p} = \frac{\partial(\tau_{\mathbf{x}_j}\sigma_j)}{\partial u_p} = \tau_{\mathbf{x}_j}\left(\frac{\partial\sigma_j}{\partial u_p}\right);$$

hence

$$\frac{\partial R}{\partial u_p} = \left(D, \tau_{\mathbf{x}_j}\left(\frac{\partial\sigma_j}{\partial u_p}\right)\right).$$

By the identity of Section 1.3.2.4.3.5, this is identical to Lifchitz's expression $(D * \partial\sigma_j/\partial u_p)(\mathbf{x}_j)$. The present derivation in terms of scalar products [see Brünger (1989) for another presentation of it] is conceptually simpler, since it invokes only the chain rule [other uses of which have been reviewed by Lunin (1985)] and Parseval's theorem; economy of computation is obviously related to the good localization of $\partial\rho^{\rm calc}/\partial u_p$ compared to $\partial F^{\rm calc}/\partial u_p$. Convolutions, whose meaning is less clear, are no longer involved; they were a legacy of having first gone over to reciprocal space *via* differential syntheses in the 1940s.

Cast in this form, the calculation of derivatives by FFT methods appears as a particular instance of the procedure described in connection with variational techniques (Section 1.3.4.4.6) to calculate the coefficients of local quadratic models in a search subspace; this is far from surprising since varying the electron density through a variation of the parameters of an atomic model is a particular case of the 'free' variations considered by the variational approach. The latter procedure would accommodate in a very natural fashion the joint consideration of an energetic (Jack & Levitt, 1978; Brünger *et al.*, 1987; Brünger, 1988; Brünger *et al.*, 1989; Kuriyan *et al.*, 1989) or stereochemical (Konnert, 1976; Sussman *et al.*, 1977; Konnert & Hendrickson, 1980; Hendrickson & Konnert, 1980; Tronrud *et al.*, 1987) restraint function (which would play the role of S) and of the crystallographic residual (which would be C). It would even have over the latter the superiority of affording a genuine *second-order* approximation, albeit only in a subspace, hence the ability of detecting negative curvature and the resulting bifurcation behaviour (Bricogne, 1984). Current methods are unable to do this because they use only *first-order* models, and this is known to degrade severely the overall efficiency of the refinement process.

1.3.4.4.7.9. Discussion of macromolecular refinement techniques

The impossibility of carrying out a full-matrix least-squares refinement of a macromolecular crystal structure, caused by excessive computational cost and by the paucity of observations, led Diamond (1971) to propose a real-space refinement method in which stereochemical knowledge was used to keep the number of free parameters to a minimum. Refinement took place by a least-squares fit between the 'observed' electron-density map and a model density consisting of Gaussian atoms. This procedure, coupled to iterative recalculation of the phases, led to the first highly refined protein structures obtained without using full-matrix least squares (Huber *et al.*, 1974; Bode & Schwager, 1975; Deisenhofer & Steigemann, 1975; Takano, 1977*a*,*b*).

Real-space refinement takes advantage of the localization of atoms (each parameter interacts only with the density near the atom to which it belongs) and gives the most immediate description of stereochemical constraints. A disadvantage is that fitting the 'observed' electron density amounts to treating the phases of the structure factors as observed quantities, and to ignoring the experimental error estimates on their moduli. The method is also much more vulnerable to series-termination errors and accidentally missing data than the least-squares method. These objections led to the progressive disuse of Diamond's method, and to a switch towards reciprocal-space least squares following Agarwal's work.

The connection established above between the Cruickshank–Agarwal modified Fourier method and the simple use of the chain rule affords a partial refutation to both the premises of Diamond's method and to the objections made against it:

(i) it shows that refinement *can* be performed through localized computations in real space *without* having to treat the phases as observed quantities;

(ii) at the same time, it shows that measurement errors on the moduli *can* be fully utilized in real space, *via* the Fourier synthesis of the functional derivative $\partial R/\partial\rho^{\rm calc}(\mathbf{x})$ or by means of the coefficients of a quadratic model of R in a search subspace.

1.3.4.4.7.10. Sampling considerations

The calculation of the inner products $(D, \partial\rho^{\rm calc}/\partial u_p)$ from a *sampled* gradient map D requires even more caution than that of structure factors *via* electron-density maps described in Section 1.3.4.4.5, because the functions $\partial\sigma_j/\partial u_p$ have transforms which extend even further in reciprocal space than the σ_j themselves. Analytically, if the σ_j are Gaussians, the $\partial\sigma_j/\partial u_p$ are finite sums of multivariate Hermite functions (Section 1.3.2.4.4.2) and hence the same is true of their transforms. The difference map D must therefore be finely sampled and the relation between error and sampling rate may be investigated as in Section 1.3.4.4.5. An examination of the sampling rates commonly used (*e.g.* one third of the resolution) shows that they are insufficient. Tronrud *et al.*

(1987) propose to relax this requirement by applying an artificial temperature factor to σ_j (*cf.* Section 1.3.4.4.5) and the negative of that temperature factor to D, a procedure of questionable validity because the latter 'sharpening' operation is ill defined [the function $\exp(\|\mathbf{x}\|^2)$ does not define a tempered distribution, so the associativity properties of convolution may be lost]. A more robust procedure would be to compute the scalar product by means of a more sophisticated numerical quadrature formula than a mere grid sum.

1.3.4.4.8. Miscellaneous correlation functions

Certain correlation functions can be useful to detect the presence of multiple copies of the same molecule (known or unknown) in the asymmetric unit of a crystal of unknown structure.

Suppose that a crystal contains one or several copies of a molecule $\mathscr{M}$ in its asymmetric unit. If $\mu(\mathbf{x})$ is the electron density of that molecule in some reference position and orientation, then

$$\mathscr{P}^0 = \sum_{j\in J}\left[\sum_{g\in G} S_g^{\#}(T_j^{\#}\mu)\right],$$

where $T_j : \mathbf{x} \longmapsto \mathbf{C}_j\mathbf{x} + \mathbf{d}_j$ describes the placement of the jth copy of the molecule with respect to the reference copy. It is assumed that each such copy is in a general position, so that there is no isotropy subgroup.

The methods of Section 1.3.4.2.2.9 (with $\mathscr{P}_j$ replaced by $C_j^{\#}\mu$, and $\mathbf{x}_j$ by $\mathbf{d}_j$) lead to the following expression for the autocorrelation of $\mathscr{P}^0$:

$$\check{\mathscr{P}}^0 * \mathscr{P}^0 = \sum_{j_1}\sum_{j_2}\sum_{g_1}\sum_{g_2} \boldsymbol{\tau}_{S_{g_2}(\mathbf{d}_{j_2})-S_{g_1}(\mathbf{d}_{j_1})} \times [(R_{g_1}^{\#}C_{j_1}^{\#}\check{\mu}) * (R_{g_2}^{\#}C_{j_2}^{\#}\mu)].$$

If μ is unknown, consider the subfamily σ of terms with $j_1 = j_2 = j$ and $g_1 = g_2 = g$:

$$\sigma = \sum_j\sum_g R_g^{\#}C_j^{\#}(\check{\mu} * \mu).$$

The scalar product $(\sigma, R^{\#}\sigma)$ in which R is a variable rotation will have a peak whenever

$$R = (R_{g_1}C_{j_1})^{-1}(R_{g_2}C_{j_2})$$

since two copies of the 'self-Patterson' $\check{\mu} * \mu$ of the molecule will be brought into coincidence. If the interference from terms in the Patterson $\pi = r * \check{\mathscr{P}}^0 * \mathscr{P}^0$ other than those present in σ is not too serious, the 'self-rotation function' $(\pi, R^{\#}\pi)$ (Rossmann & Blow, 1962; Crowther, 1972) will show the same peaks, from which the rotations $\{C_j\}_{j\in J}$ may be determined, either individually or jointly if for instance they form a group.

If μ is known, then its self-Patterson $\check{\mu} * \mu$ may be calculated, and the C_j may be found by examining the 'cross-rotation function' $[\pi, R^{\#}(\check{\mu} * \mu)]$ which will have peaks at $R = R_gC_j$, $g \in G, j \in J$. Once the C_j are known, then the various copies $C_j^{\#}\mu$ of $\mathscr{M}$ may be Fourier-analysed into structure factors:

$$M_j(\mathbf{h}) = \bar{\mathscr{F}}[C_j^{\#}\mu](\mathbf{h}).$$

The cross terms with $j_1 \neq j_2, g_1 \neq g_2$ in $\check{\mathscr{P}}^0 * \mathscr{P}^0$ then contain 'motifs'

$$(R_{g_1}^{\#}C_{j_1}^{\#}\check{\mu}) * (R_{g_2}^{\#}C_{j_2}^{\#}\mu),$$

with Fourier coefficients

$$\overline{M_{j_1}(\mathbf{R}_{g_1}^T\mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T\mathbf{h}),$$

translated by $S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$. Therefore the 'translation functions' (Crowther & Blow, 1967)

$$\mathscr{T}_{j_1g_1,j_2g_2}(\mathbf{s}) = \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \overline{M_{j_1}(\mathbf{R}_{g_1}^T\mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T\mathbf{h}) \exp(-2\pi i\mathbf{h}\cdot\mathbf{s})$$

will have peaks at $\mathbf{s} = S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$ corresponding to the detection of these motifs.

1.3.4.5. Related applications

1.3.4.5.1. Helical diffraction

The theory of diffraction by helical structures (Cochran *et al.*, 1952; Klug *et al.*, 1958) has played an important part in the study of polypeptides, of nucleic acids and of tobacco mosaic virus.

1.3.4.5.1.1. Circular harmonic expansions in polar coordinates

Let $f = f(x, y)$ be a reasonably regular function in two-dimensional real space. Going over to polar coordinates

$$x = r\cos\varphi \quad y = r\sin\varphi$$

and writing, by slight misuse of notation, $f(r, \varphi)$ for $f(r\cos\varphi, r\sin\varphi)$ we may use the periodicity of f with respect to φ to expand it as a Fourier series (Byerly, 1893):

$$f(r, \varphi) = \sum_{n\in\mathbb{Z}} f_n(r)\exp(in\varphi)$$

with

$$f_n(r) = \frac{1}{2\pi}\int_0^{2\pi} f(r, \varphi)\exp(-in\varphi)\ \mathrm{d}\varphi.$$

Similarly, in reciprocal space, if $F = F(\xi, \eta)$ and if

$$\xi = R\cos\psi \quad \eta = R\sin\psi$$

then

$$F(R, \psi) = \sum_{n\in\mathbb{Z}} i^n F_n(R)\exp(in\psi)$$

with

$$F_n(R) = \frac{1}{2\pi i^n}\int_0^{2\pi} F(R, \psi)\exp(-in\psi)\ \mathrm{d}\psi,$$

where the phase factor i^n has been introduced for convenience in the forthcoming step.

1.3.4.5.1.2. The Fourier transform in polar coordinates

The Fourier transform relation between f and F may then be written in terms of f_n's and F_n's. Observing that $\xi x + \eta y = Rr\cos(\varphi - \psi)$, and that (Watson, 1944)

$$\int_0^{2\pi} \exp(iX\cos\theta + in\theta)\ \mathrm{d}\theta = 2\pi i^n J_n(X),$$

we obtain:

$$\begin{aligned} F(R,\psi) &= \int_0^{\infty}\int_0^{2\pi}\left[\sum_{n\in\mathbb{Z}} f_n(r)\exp(in\varphi)\right] \\ &\quad\times \exp[2\pi iRr\cos(\varphi-\psi)]r\ \mathrm{d}r\ \mathrm{d}\varphi \\ &= \sum_{n\in\mathbb{Z}} i^n\left[\int_0^{\infty} f_n(r)J_n(2\pi Rr)2\pi r\ \mathrm{d}r\right]\exp(in\psi); \end{aligned}$$

hence, by the uniqueness of the Fourier expansion of F:

$$F_n(R) = \int_0^{\infty} f_n(r)J_n(2\pi Rr)2\pi r\ \mathrm{d}r.$$

The inverse Fourier relationship leads to

$$f_n(r) = \int_0^{\infty} F_n(R)J_n(2\pi rR)2\pi R\ \mathrm{d}R.$$

The integral transform involved in the previous two equations is called the *Hankel transform* (see *e.g.* Titchmarsh, 1922; Sneddon, 1972) of order n.

1.3.4.5.1.3. The transform of an axially periodic fibre

Let ρ be the electron-density distribution in a fibre, which is assumed to have translational periodicity with period 1 along z, and to have compact support with respect to the (x, y) coordinates. Thus ρ may be written

$$\rho = \left[\delta_x \otimes \delta_y \otimes \left(\sum_{k\in\mathbb{Z}}\delta_{(k)}\right)_z\right] * \rho^0,$$

where $\rho^0 = \rho^0(x, y, z)$ is the motif.

By the tensor product property, the inverse Fourier transform $F = \bar{\mathscr{F}}_{xyz}[\rho]$ may be written

$$F = \left[1_\xi \otimes 1_\eta \otimes \left(\sum_{l\in\mathbb{Z}}\delta_{(l)}\right)_\zeta\right] \times \bar{\mathscr{F}}[\rho^0]$$

and hence consists of 'layers' labelled by l:

$$F = \sum_{l\in\mathbb{Z}} F(\xi,\eta,l)(\delta_{(l)})_\zeta$$

with

$$F(\xi,\eta,l) = \int_0^1 \bar{\mathscr{F}}_{xy}[\rho^0](\xi,\eta,z)\exp(2\pi ilz)\ \mathrm{d}z.$$

Changing to polar coordinates in the (x, y) and (ξ, η) planes decomposes the calculation of F from ρ into the following steps:

$$\begin{aligned} g_{nl}(r) &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \rho(r,\varphi,z)\exp[i(-n\varphi + 2\pi lz)]\ \mathrm{d}\varphi\ \mathrm{d}z \\ G_{nl}(R) &= \int_0^{\infty} g_{nl}(r)J_n(2\pi Rr)2\pi r\ \mathrm{d}r \\ F(R,\psi,l) &= \sum_{n\in\mathbb{Z}} i^n G_{nl}(R)\exp(in\psi) \end{aligned}$$

and the calculation of ρ from F into:

$$\begin{aligned} G_{nl}(R) &= \frac{1}{2\pi i^n}\int_0^{2\pi} F(R,\psi,l)\exp(-in\psi)\ \mathrm{d}\psi \\ g_{nl}(r) &= \int_0^{\infty} G_{nl}(R)J_n(2\pi rR)2\pi R\ \mathrm{d}R \\ \rho(r,\varphi,z) &= \sum_{n\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} g_{nl}(r)\exp[i(n\varphi - 2\pi lz)]. \end{aligned}$$

These formulae are seen to involve a 2D Fourier series with respect to the two periodic coordinates φ and z, and Hankel transforms along the radial coordinates. The two periodicities in φ and z are independent, so that all combinations of indices (n, l) occur in the Fourier summations.

1.3.4.5.1.4. Helical symmetry and associated selection rules

Helical symmetry involves a 'clutching' between the two (hitherto independent) periodicities in φ (period 2π) and z (period 1) which causes a subdivision of the period lattice and hence a decimation (governed by 'selection rules') of the Fourier coefficients.

Let **i** and **j** be the basis vectors along $\varphi/2\pi$ and z. The integer lattice with basis **(i, j)** is a period lattice for the (φ, z) dependence of the electron density ρ of an axially periodic fibre considered in Section 1.3.4.5.1.3:

$$\rho(r, \varphi + 2\pi k_1, z + k_2) = \rho(r,\varphi,z).$$

Suppose the fibre now has *helical* symmetry, with u copies of the same molecule in t turns, where g.c.d. $(u, t) = 1$. Using the Euclidean algorithm, write $u = \lambda t + \mu$ with λ and μ positive integers and $\mu < t$. The period lattice for the (φ, z) dependence of ρ may be defined in terms of the new basis vectors:

I, joining subunit 0 to subunit l in the same turn;

J, joining subunit 0 to subunit λ after wrapping around.

In terms of the original basis

$$\mathbf{I} = \frac{t}{u}\mathbf{i} + \frac{1}{u}\mathbf{j}, \quad \mathbf{J} = \frac{-\mu}{u}\mathbf{i} + \frac{\lambda}{u}\mathbf{j}.$$

If α and β are coordinates along **I** and **J**, respectively,

$$\begin{pmatrix}\varphi/2\pi\\ z\end{pmatrix} = \frac{1}{u}\begin{pmatrix}t & -\mu\\ 1 & \lambda\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

or equivalently

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}\lambda & \mu\\ -1 & t\end{pmatrix}\begin{pmatrix}\varphi/2\pi\\ z\end{pmatrix}.$$

By Fourier transformation,

$$\left(\frac{\varphi}{2\pi}, z\right) \Leftrightarrow (-n, l)$$
$$(\alpha, \beta) \Leftrightarrow (m, p)$$

with the transformations between indices given by the contragredients of those between coordinates, *i.e.*

$$\begin{pmatrix} n \\ l \end{pmatrix} = \begin{pmatrix} -\lambda & 1 \\ -\mu & t \end{pmatrix} \begin{pmatrix} m \\ p \end{pmatrix}$$

and

$$\begin{pmatrix} m \\ p \end{pmatrix} = \frac{1}{u} \begin{pmatrix} -t & 1 \\ \mu & \lambda \end{pmatrix} \begin{pmatrix} n \\ l \end{pmatrix}.$$

It follows that

$$l = tn + um,$$

or alternatively that

$$\mu n = up - \lambda l,$$

which are two equivalent expressions of the *selection rules* describing the decimation of the transform. These rules imply that only certain orders n contribute to a given layer l.

The 2D Fourier analysis may now be performed by analysing a *single* subunit referred to coordinates α and β to obtain

$$h_{m,p}(r) = \int_0^1 \int_0^1 \rho(r, \alpha, \beta) \exp[2\pi i(m\alpha + p\beta)] \, \mathrm{d}\alpha \, \mathrm{d}\beta$$

and then reindexing to get only the allowed g_{nl}'s by

$$g_{nl}(r) = u h_{-\lambda m + p, \mu m + tp}(r).$$

This is u times faster than analysing u subunits with respect to the (φ, z) coordinates.

1.3.4.5.2. Application to probability theory and direct methods

The Fourier transformation plays a central role in the branch of probability theory concerned with the limiting behaviour of sums of large numbers of independent and identically distributed random variables or random vectors. This privileged role is a consequence of the convolution theorem and of the 'moment-generating' properties which follow from the exchange between differentiation and multiplication by monomials. When the limit theorems are applied to the calculation of joint probability distributions of structure factors, which are themselves closely related to the Fourier transformation, a remarkable phenomenon occurs, which leads to the saddlepoint approximation and to the maximum-entropy method.

1.3.4.5.2.1. Analytical methods of probability theory

The material in this section is not intended as an introduction to probability theory [for which the reader is referred to Cramér (1946), Petrov (1975) or Bhattacharya & Rao (1976)], but only as an illustration of the role played by the Fourier transformation in certain specific areas which are used in formulating and implementing direct methods of phase determination.

(*a*) *Convolution of probability densities*

The addition of independent random variables or vectors leads to the convolution of their probability distributions: if $\mathbf{X}_1$ and $\mathbf{X}_2$ are two n-dimensional random vectors independently distributed with probability densities P_1 and P_2, respectively, then their sum $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ has probability density $\mathscr{P}$ given by

$$\begin{aligned} \mathscr{P}(\mathbf{X}) &= \int_{\mathbf{R}^n} P_1(\mathbf{X}_1) P_2(\mathbf{X} - \mathbf{X}_1) \, \mathrm{d}^n\mathbf{X}_1 \\ &= \int_{\mathbf{R}^n} P_1(\mathbf{X} - \mathbf{X}_2) P_2(\mathbf{X}_2) \, \mathrm{d}^n\mathbf{X}_2 \end{aligned}$$

i.e.

$$\mathscr{P} = P_1 * P_2.$$

This result can be extended to the case where P_1 and P_2 are singular measures (distributions of order zero, Section 1.3.2.3.4) and do not have a density with respect to the Lebesgue measure in $\mathbb{R}^n$.

(*b*) *Characteristic functions*

This convolution can be turned into a simple multiplication by considering the Fourier transforms (called the *characteristic functions*) of P_1, P_2 and $\mathscr{P}$, defined with a slightly different normalization in that there is no factor of 2π in the exponent (see Section 1.3.2.4.5), *e.g.*

$$C(\mathbf{t}) = \int_{\mathbf{R}^n} P(\mathbf{X}) \exp(i\mathbf{t} \cdot \mathbf{X}) \, \mathrm{d}^n\mathbf{X}.$$

Then by the convolution theorem

$$\mathscr{C}(\mathbf{t}) = C_1(\mathbf{t}) \times C_2(\mathbf{t}),$$

so that $\mathscr{P}(\mathbf{X})$ may be evaluated by Fourier inversion of its characteristic function as

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} C_1(\mathbf{t}) C_2(\mathbf{t}) \exp(-i\mathbf{t} \cdot \mathbf{X}) \, \mathrm{d}^n\mathbf{t}$$

(see Section 1.3.2.4.5 for the normalization factors).

It follows from the differentiation theorem that the partial derivatives of the characteristic function $C(\mathbf{t})$ at $\mathbf{t} = \mathbf{0}$ are related to the moments of a distribution $\boldsymbol{P}$ by the identities

$$\begin{aligned} \mu_{r_1 r_2 \ldots r_n} &\equiv \int_D P(\mathbf{X}) X_1^{r_1} X_2^{r_2} \ldots X_n^{r_n} \, \mathrm{d}^n\mathbf{X} \\ &= i^{-(r_1 + \ldots + r_n)} \frac{\partial^{r_1 + \ldots + r_n} C}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}} \bigg|_{\mathbf{t}=\mathbf{0}} \end{aligned}$$

for any n-tuple of non-negative integers $(r_1, r_2, \ldots, r_n)$.

(*c*) *Moment-generating functions*

The above relation can be freed from powers of i by defining (at least formally) the *moment-generating function*:

$$M(\mathbf{t}) = \int_{\mathbb{R}^n} P(\mathbf{X}) \exp(\mathbf{t} \cdot \mathbf{X}) \, \mathrm{d}^n\mathbf{X}$$

which is related to $C(\mathbf{t})$ by $C(\mathbf{t}) = \boldsymbol{M}(i\mathbf{t})$ so that the inversion formula reads

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} M_1(i\mathbf{t}) M_2(i\mathbf{t}) \exp(-i\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{t}.$$

The moment-generating function is well defined, in particular, for any probability distribution with compact support, in which case it may be continued analytically from a function over $\mathbb{R}^n$ into an entire function of *n complex* variables by virtue of the Paley–Wiener theorem (Section 1.3.2.4.2.10). Its moment-generating properties are summed up in the following relations:

$$\mu_{r_1 r_2 \ldots r_n} = \left. \frac{\partial^{r_1+\ldots+r_n} M}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}} \right|_{\mathbf{t}=\mathbf{0}}.$$

(*d*) *Cumulant-generating functions*

The multiplication of moment-generating functions may be further simplified into the addition of their logarithms:

$$\log \mathscr{M} = \log M_1 + \log M_2,$$

or equivalently of the coefficients of their Taylor series at $\mathbf{t} = \mathbf{0}$, *viz*:

$$\kappa_{r_1 r_2 \ldots r_n} = \left. \frac{\partial^{r_1+\ldots+r_n} (\log M)}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}} \right|_{\mathbf{t}=\mathbf{0}}.$$

These coefficients are called *cumulants*, since they add when the independent random vectors to which they belong are added, and $\log M$ is called the *cumulant-generating function.* The inversion formula for $\mathscr{P}$ then reads

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[\log M_1(i\mathbf{t}) + \log M_2(i\mathbf{t}) - i\mathbf{t}\cdot\mathbf{X}]\ \mathrm{d}^n\mathbf{t}.$$

(*e*) *Asymptotic expansions and limit theorems*

Consider an n-dimensional random vector $\mathbf{X}$ of the form

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_N,$$

where the N summands are independent n-dimensional random vectors identically distributed with probability density P. Then the distribution $\mathscr{P}$ of $\mathbf{X}$ may be written in closed form as a Fourier transform:

$$\begin{aligned}\mathscr{P}(\mathbf{X}) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} M^N(i\mathbf{t}) \exp(-i\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{t} \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[N \log M(i\mathbf{t}) - i\mathbf{t}\cdot\mathbf{X}]\ \mathrm{d}^n\mathbf{t},\end{aligned}$$

where

$$M(\mathbf{t}) = \int_{\mathbb{R}^n} P(\mathbf{Y}) \exp(\mathbf{t}\cdot\mathbf{Y})\ \mathrm{d}^n\mathbf{Y}$$

is the moment-generating function common to all the summands.

This an exact expression for $\mathscr{P}$, which may be exploited analytically or numerically in certain favourable cases. Supposing for instance that P has compact support, then its characteristic function $M(i\mathbf{t})$ can be sampled finely enough to accommodate the bandwidth of the support of $\mathscr{P} = P^{*N}$ (this sampling rate clearly depends on n) so that the above expression for $\mathscr{P}$ can be used for its numerical evaluation as the *discrete* Fourier transform of $M^N(i\mathbf{t})$. This exact method is practical only for small values of the dimension n.

In all other cases some form of approximation must be used in the Fourier inversion of $M^N(i\mathbf{t})$. For this purpose it is customary (Cramér, 1946) to expand the cumulant-generating function around $\mathbf{t} = \mathbf{0}$ with respect to the carrying variables $\mathbf{t}$:

$$\log[M^N(i\mathbf{t})] = \sum_{\mathbf{r}\in\mathbb{N}^n} \frac{N\kappa_{\mathbf{r}}}{\mathbf{r}!} (i\mathbf{t})^{\mathbf{r}},$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ is a multi-index (Section 1.3.2.2.3). The first-order terms may be eliminated by recentring $\mathscr{P}$ around its vector of first-order cumulants

$$\langle \mathbf{X} \rangle = \sum_{j=1}^{N} \langle \mathbf{X}_j \rangle,$$

where $\langle\cdot\rangle$ denotes the mathematical expectation of a random vector. The second-order terms may be grouped separately from the terms of third or higher order to give

$$\begin{aligned}M^N(i\mathbf{t}) = \exp(-\tfrac{1}{2} N \mathbf{t}^U \mathbf{Q} \mathbf{t}) \\ \times \exp\left\{ \sum_{|\mathbf{r}|\geq 3} \frac{N\kappa_{\mathbf{r}}}{\mathbf{r}!} (i\mathbf{t})^{\mathbf{r}} \right\},\end{aligned}$$

where $\mathbf{Q} = \nabla\nabla^T (\log M)$ is the covariance matrix of the multivariate distribution P. Expanding the exponential gives rise to a series of terms of the form

$$\exp(-\tfrac{1}{2} N \mathbf{t}^T \mathbf{Q} \mathbf{t}) \times \text{monomial in } t_1, t_2, \ldots, t_n,$$

each of which may now be subjected to a Fourier transformation to yield a Hermite function of $\mathbf{t}$ (Section 1.3.2.4.4.2) with coefficients involving the cumulants κ of P. Taking the transformed terms in natural order gives an asymptotic expansion of P for large N called the *Gram–Charlier series* of $\mathscr{P}$, while grouping the terms according to increasing powers of $1/\sqrt{N}$ gives another asymptotic expansion called the *Edgeworth series* of $\mathscr{P}$. Both expansions comprise a leading Gaussian term which embodies the *central-limit theorem*:

$$\mathscr{P}(\mathbf{E}) = \frac{1}{\sqrt{\det(2\pi\mathbf{Q})}} \exp(-\tfrac{1}{2}\mathbf{E}^T \mathbf{Q}^{-1} \mathbf{E}), \quad \text{where } \mathbf{E} = \frac{\mathbf{X} - \langle \mathbf{X} \rangle}{\sqrt{N}}.$$

(*f*) *The saddlepoint approximation*

A limitation of the Edgeworth series is that it gives an accurate estimate of $\mathscr{P}(\mathbf{X})$ only in the vicinity of $\mathbf{X} = \langle \mathbf{X} \rangle$, *i.e.* for small values of $\mathbf{E}$. These convergence difficulties are easily understood: one is substituting a *local* approximation to $\log M$ (*viz* a Taylor-series expansion valid near $\mathbf{t} = \mathbf{0}$) into an integral, whereas integration is a *global* process which consults values of $\log M$ far from $\mathbf{t} = \mathbf{0}$.

It is possible, however, to let the point $\mathbf{t}$ where $\log M$ is expanded as a Taylor series depend on the particular value $\mathbf{X}^*$ of $\mathbf{X}$ for which an accurate evaluation of $\mathscr{P}(\mathbf{X})$ is desired. This is the essence of the *saddlepoint method* (Fowler, 1936; Khinchin 1949;

Daniels, 1954; de Bruijn, 1970; Bleistein & Handelsman, 1986), which uses an analytical continuation of $M(\mathbf{t})$ from a function over $\mathbb{R}^n$ to a function over $\mathbb{C}^n$ (see Section 1.3.2.4.2.10). Putting then $\mathbf{t} = \mathbf{s} - i\tau$, the $\mathbb{C}^n$ version of Cauchy's theorem (Hörmander, 1973) gives rise to the identity

$$\mathscr{P}(\mathbf{X}^*) = \frac{\exp(-\boldsymbol{\tau}\cdot\mathbf{X}^*)}{(2\pi)^n} \times \int_{\mathbb{R}^n} \exp\left\{N\left[\log M(\boldsymbol{\tau}+i\mathbf{s}) - i\mathbf{s}\cdot\frac{\mathbf{X}^*}{N}\right]\right\} \mathrm{d}^n\mathbf{s}$$

for *any* $\boldsymbol{\tau}\in\mathbb{R}^n$. By a convexity argument involving the positive-definiteness of covariance matrix $\mathbf{Q}$, there is a unique value of $\boldsymbol{\tau}$ such that

$$\nabla(\log M)|_{\mathbf{t}=\mathbf{0}-i\boldsymbol{\tau}} = \frac{\mathbf{X}^*}{N}.$$

At the *saddlepoint* $\mathbf{t}^* = \mathbf{0} - i\boldsymbol{\tau}$, the modulus of the integrand above is a maximum and its phase is stationary with respect to the integration variable $\mathbf{s}$: as N tends to infinity, all contributions to the integral cancel because of rapid oscillation, except those coming from the immediate vicinity of $\mathbf{t}^*$ where there is no oscillation. A Taylor expansion of $\log M^N$ to second order with respect to $\mathbf{s}$ at $\mathbf{t}^*$ then gives

$$\log M^N(\boldsymbol{\tau}+i\mathbf{s}) \approx \log M^N(\boldsymbol{\tau}) + i\mathbf{s}\cdot\mathbf{X}^* - \frac{N}{2}[\mathbf{s}^T\mathbf{Q}\mathbf{s}]$$

and hence

$$\mathscr{P}(\mathbf{X}^*) \approx \exp[\log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau}\cdot\mathbf{X}^*]\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\mathbf{s}^T\mathscr{Q}\mathbf{s})\ \mathrm{d}^n\mathbf{s}.$$

The last integral is elementary and gives the 'saddlepoint approximation':

$$\mathscr{P}^{\mathrm{SP}}(\mathbf{X}^*) = \frac{\exp(\mathsf{S})}{\sqrt{\det(2\pi\mathscr{Q})}},$$

where

$$\mathsf{S} = \log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau}\cdot\mathbf{X}^*$$

and where

$$\mathscr{Q} = \nabla\nabla^T(\log M^N) = N\mathbf{Q}.$$

This approximation scheme amounts to using the 'conjugate distribution' (Khinchin, 1949)

$$P_{\boldsymbol{\tau}}(\mathbf{X}_j) = P(\mathbf{X}_j)\frac{\exp(\boldsymbol{\tau}\cdot\mathbf{X}_j)}{M(\boldsymbol{\tau})}$$

instead of the original distribution $\boldsymbol{P}(\mathbf{X}_j) = \boldsymbol{P}_{\mathbf{0}}(\mathbf{X}_j)$ for the common distribution of all $\boldsymbol{N}$ random vectors $\mathbf{X}_j$. The exponential modulation results from the analytic continuation of the characteristic (or moment-generating) function into $\mathbb{C}^n$, as in Section 1.3.2.4.2.10. The saddlepoint approximation $\mathscr{P}^{\mathrm{SP}}$ is only the leading term of an asymptotic expansion (called the *saddlepoint expansion*) for $\mathscr{P}$, which is actually the Edgeworth expansion associated with $P_{\boldsymbol{\tau}}^{*N}$.

1.3.4.5.2.2. The statistical theory of phase determination

The methods of probability theory just surveyed were applied to various problems formally similar to the crystallographic phase problem [*e.g.* the 'problem of the random walk' of Pearson (1905)] by Rayleigh (1880, 1899, 1905, 1918, 1919) and Kluyver (1906). They became the basis of the statistical theory of communication with the classic papers of Rice (1944, 1945).

The Gram–Charlier and Edgeworth series were introduced into crystallography by Bertaut (1955*a*,*b*,*c*, 1956*a*) and by Klug (1958), respectively, who showed them to constitute the mathematical basis of numerous formulae derived by Hauptman & Karle (1953). The saddlepoint approximation was introduced by Bricogne (1984) and was shown to be related to variational methods involving the maximization of certain entropy criteria. This connection exhibits most of the properties of the Fourier transform at play simultaneously, and will now be described as a final illustration.

(*a*) *Definitions and conventions*

Let H be a set of unique non-origin reflections $\mathbf{h}$ for a crystal with lattice Λ and space group G. Let H contain n_{a} acentric and n_{c} centric reflections. Structure-factor values attached to all reflections in H will comprise $n = 2n_{\mathrm{a}} + n_{\mathrm{c}}$ real numbers. For $\mathbf{h}$ acentric, $\alpha_{\mathbf{h}}$ and $\beta_{\mathbf{h}}$ will be the real and imaginary parts of the complex structure factor; for $\mathbf{h}$ centric, $\gamma_{\mathbf{h}}$ will be the real coordinate of the (possibly complex) structure factor measured along a real axis rotated by one of the two angles $\theta_{\mathbf{h}}$, π apart, to which the phase is restricted modulo 2π (Section 1.3.4.2.2.5). These n real coordinates will be arranged as a column vector containing the acentric then the centric data, *i.e.* in the order

$$\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_{n_{\mathrm{a}}}, \beta_{n_{\mathrm{a}}}, \gamma_1, \gamma_2, \ldots, \gamma_{n_{\mathrm{c}}}.$$

(*b*) *Vectors of trigonometric structure-factor expressions*

Let $\boldsymbol{\xi}(\mathbf{x})$ denote the vector of trigonometric structure-factor expressions associated with $\mathbf{x}\in D$, where D denotes the asymmetric unit. These are defined as follows:

$$\begin{aligned}\alpha_{\mathbf{h}}(\mathbf{x}) + i\beta_{\mathbf{h}}(\mathbf{x}) &= \Xi(\mathbf{h},\mathbf{x}) \quad \text{for } \mathbf{h} \text{ acentric}\\ \gamma_{\mathbf{h}}(\mathbf{x}) &= \exp(-i\theta_{\mathbf{h}})\Xi(\mathbf{h},\mathbf{x}) \quad \text{for } \mathbf{h} \text{ centric,}\end{aligned}$$

where

$$\Xi(\mathbf{h},\mathbf{x}) = \frac{1}{|G_{\mathbf{x}}|}\sum_{g\in G}\exp\{2\pi i\mathbf{h}\cdot[S_g(\mathbf{x})]\}.$$

According to the convention above, the coordinates of $\boldsymbol{\xi}(\mathbf{x})$ in $\mathbb{R}^n$ will be arranged in a column vector as follows:

$$\begin{aligned}\xi_{2r-1}(\mathbf{x}) &= \alpha_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = 1,\ldots,n_{\mathrm{a}},\\ \xi_{2r}(\mathbf{x}) &= \beta_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = 1,\ldots,n_{\mathrm{a}},\\ \xi_{n_{\mathrm{a}}+r}(\mathbf{x}) &= \gamma_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = n_{\mathrm{a}}+1,\ldots,n_{\mathrm{a}}+n_{\mathrm{c}}.\end{aligned}$$

(*c*) *Distributions of random atoms and moment-generating functions*

Let position $\mathbf{x}$ in D now become a random vector with probability density $m(\mathbf{x})$. Then $\boldsymbol{\xi}(\mathbf{x})$ becomes itself a random vector in $\mathbb{R}^n$, whose distribution $p(\boldsymbol{\xi})$ is the image of distribution $m(\mathbf{x})$

through the mapping $\mathbf{x} \rightarrow \boldsymbol{\xi}(\mathbf{x})$ just defined. The locus of $\boldsymbol{\xi}(\mathbf{x})$ in $\mathbb{R}^n$ is a compact algebraic manifold $\mathscr{V}$ (the multidimensional analogue of a Lissajous curve), so that p is a singular measure (a distribution of order 0, Section 1.3.2.3.4, concentrated on that manifold) with compact support. The average with respect to p of any function Ω over $\mathbb{R}^n$ which is infinitely differentiable in a neighbourhood of $\mathscr{V}$ may be calculated as an average with respect to m over D by the 'induction formula':

$$\langle p, \Omega \rangle = \int_D m(\mathbf{x}) \Omega[\boldsymbol{\xi}(\mathbf{x})] \, \mathrm{d}^3\mathbf{x}.$$

In particular, one can calculate the moment-generating function M for distribution p as

$$M(\mathbf{t}) \equiv \langle p_\xi, \exp(\mathbf{t} \cdot \boldsymbol{\xi}) \rangle = \int_D m(\mathbf{x}) \exp[\mathbf{t} \cdot \boldsymbol{\xi}(\mathbf{x})] \, \mathrm{d}^3\mathbf{x}$$

and hence calculate the moments μ (respectively cumulants κ) of p by differentiation of M (respectively $\log M$) at $\mathbf{t} = \mathbf{0}$:

$$\begin{aligned} \mu_{r_1 r_2 \ldots r_n} &\equiv \int_D m(\mathbf{x}) \xi_1^{r_1}(\mathbf{x}) \xi_2^{r_2}(\mathbf{x}) \ldots \xi_n^{r_n}(\mathbf{x}) \, \mathrm{d}^3\mathbf{x} \\ &= \frac{\partial^{r_1+\ldots+r_n}(M)}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}} \\ \kappa_{r_1 r_2 \ldots r_n} &= \frac{\partial^{r_1+\ldots+r_n}(\log M)}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}}. \end{aligned}$$

The structure-factor algebra for group G (Section 1.3.4.2.2.9) then allows one to express products of $\boldsymbol{\xi}$'s as linear combinations of other $\boldsymbol{\xi}$'s, and hence to express all moments and cumulants of distribution $p(\boldsymbol{\xi})$ as linear combinations of real and imaginary parts of Fourier coefficients of the prior distribution of atoms $m(\mathbf{x})$. This plays a key role in the use of nonuniform distributions of atoms.

(*d*) *The joint probability distribution of structure factors*

In the random-atom model of an equal-atom structure, N atoms are placed randomly, independently of each other, in the asymmetric unit D of the crystal with probability density $m(\mathbf{x})$. For point atoms of unit weight, the vector $\mathbf{F}$ of structure-factor values for reflections $\mathbf{h} \in H$ may be written

$$\mathbf{F} = \sum_{I=1}^{N} \boldsymbol{\xi}^{[I]},$$

where the N copies $\boldsymbol{\xi}^{[I]}$ of random vector $\boldsymbol{\xi}$ are independent and have the same distribution $p(\boldsymbol{\xi})$.

The joint probability distribution $\mathscr{P}(\mathbf{F})$ is then [Section 1.3.4.5.2.1(*e*)]

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[N \log M(i\mathbf{t}) - i\mathbf{t} \cdot \mathbf{X}] \, \mathrm{d}^n\mathbf{t}.$$

For low dimensionality n it is possible to carry out the Fourier transformation numerically after discretization, provided $M(i\mathbf{t})$ is sampled sufficiently finely that no aliasing results from taking its Nth power (Barakat, 1974). This exact approach can also accommodate heterogeneity, and has been used first in the field of intensity statistics (Shmueli *et al.*, 1984, 1985; Shmueli & Weiss, 1987, 1988), then in the study of the Σ_1 and Σ_2 relations in triclinic space groups (Shmueli & Weiss, 1985, 1986). Some of these applications are described in Chapter 2.1 of this volume. This method could be extended to the construction of any joint probability distribution (j.p.d.) in any space group by using the generic expression for the moment-generating function (m.g.f.) derived by Bricogne (1984). It is, however, limited to small values of n by the necessity to carry out n-dimensional FFTs on large arrays of sample values.

The asymptotic expansions of Gram–Charlier and Edgeworth have good convergence properties only if $F_{\mathbf{h}}$ lies in the vicinity of $\langle F_{\mathbf{h}} \rangle = N\bar{\mathscr{F}}[m](\mathbf{h})$ for all $\mathbf{h} \in H$. Previous work on the j.p.d. of structure factors has used for $m(\mathbf{x})$ a uniform distribution, so that $\langle \mathbf{F} \rangle = \mathbf{0}$; as a result, the corresponding expansions are accurate only if all moduli $|F_{\mathbf{h}}|$ are small, in which case the j.p.d. contains little phase information.

The saddlepoint method [Section 1.3.4.5.2.1(*f*)] constitutes the method of choice for evaluating the joint probability $\mathscr{P}(\mathbf{F}^*)$ of structure factors when some of the moduli in $\mathbf{F}^*$ are large. As shown previously, this approximation amounts to using the 'conjugate distribution'

$$p_\tau(\xi) = p(\xi) \frac{\exp(\boldsymbol{\tau} \cdot \boldsymbol{\xi})}{M(\boldsymbol{\tau})}$$

instead of the original distribution $p(\boldsymbol{\xi}) = p_{\mathbf{0}}(\boldsymbol{\xi})$ for the distribution of random vector $\boldsymbol{\xi}$. This conjugate distribution $p_{\boldsymbol{\tau}}$ is induced from the modified distribution of atoms

$$q_{\boldsymbol{\tau}}(\mathbf{x}) = m(\mathbf{x}) \frac{\exp[\boldsymbol{\tau} \cdot \boldsymbol{\xi}(\mathbf{x})]}{M(\boldsymbol{\tau})}, \tag{SP1}$$

where, by the induction formula, $M(\boldsymbol{\tau})$ may be written as

$$M(\boldsymbol{\tau}) = \int_D m(\mathbf{x}) \exp[\boldsymbol{\tau} \cdot \boldsymbol{\xi}(\mathbf{x})] \, \mathrm{d}^3\mathbf{x} \tag{SP2}$$

and where $\boldsymbol{\tau}$ is the unique solution of the saddlepoint equation:

$$\nabla_{\boldsymbol{\tau}}(\log M^N) = \mathbf{F}^*. \tag{SP3}$$

The desired approximation is then

$$\mathscr{P}^{\mathrm{SP}}(\mathbf{F}^*) = \frac{\exp(\mathsf{S})}{\sqrt{\det(2\pi \mathscr{Q})}},$$

where

$$\mathsf{S} = \log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau} \cdot \mathbf{F}^*$$

and where

$$\mathscr{Q} = \nabla\nabla^T(\log M^N) = \mathbf{NQ}.$$

Finally, the elements of the Hessian matrix $\mathbf{Q} = \nabla\nabla^T(\log M)$ are just the trigonometric second-order cumulants of distribution p, and hence can be calculated *via* structure-factor algebra from the Fourier coefficients of $q_{\boldsymbol{\tau}}(\mathbf{x})$. All the quantities involved in the expression for $\mathscr{P}^{\mathrm{SP}}(\mathbf{F}^*)$ are therefore effectively computable from the initial data $m(\mathbf{x})$ and $\mathbf{F}^*$.

(*e*) *Maximum-entropy distributions of atoms*

One of the main results in Bricogne (1984) is that the modified distribution $q_{\boldsymbol{\tau}}(\mathbf{x})$ in (SP1) is the unique distribution which has maximum entropy $\mathscr{S}_m(q)$ relative to $m(\mathbf{x})$, where

$$\mathscr{S}_m(q) = -\int_D q(\mathbf{x}) \log\left[\frac{q(\mathbf{x})}{m(\mathbf{x})}\right] \mathrm{d}^3\mathbf{x},$$

under the constraint that $\mathbf{F}^*$ be the centroid vector of the corresponding conjugate distribution $\mathscr{P}_{\boldsymbol{\tau}}(\mathbf{F})$. The traditional notation of maximum-entropy (ME) theory (Jaynes, 1957, 1968, 1983) is in this case (Bricogne, 1984)

$$q^{\mathrm{ME}}(\mathbf{x}) = m(\mathbf{x}) \frac{\exp[\boldsymbol{\lambda} \cdot \boldsymbol{\xi}(\mathbf{x})]}{Z(\boldsymbol{\lambda})} \qquad \text{(ME1)}$$

$$Z(\boldsymbol{\lambda}) = \int_D m(\mathbf{x}) \exp[\boldsymbol{\lambda} \cdot \boldsymbol{\xi}(\mathbf{x})] \, \mathrm{d}^3\mathbf{x} \qquad \text{(ME2)}$$

$$\nabla_\lambda (\log Z^N) = \mathbf{F}^* \qquad \text{(ME3)}$$

so that Z is identical to the m.g.f. M, and the coordinates $\boldsymbol{\tau}$ of the saddlepoint are the Lagrange multipliers $\boldsymbol{\lambda}$ for the constraints $\mathbf{F}^*$.

Jaynes's ME theory also gives an estimate for $\mathscr{P}(\mathbf{F}^*)$:

$$\mathscr{P}^{\mathrm{ME}}(\mathbf{F}^*) \approx \exp(\mathscr{S}),$$

where

$$\mathscr{S} = \log Z^N - \boldsymbol{\lambda} \cdot \mathbf{F}^* = N \mathscr{S}_m(\mathbf{q}^{\mathrm{ME}})$$

is the total entropy and is the counterpart to S under the equivalence just established.

$\mathscr{P}^{\mathrm{ME}}$ is identical to $\mathscr{P}^{\mathrm{SP}}$, but lacks the denominator. The latter, which is the normalization factor of a multivariate Gaussian with covariance matrix $\mathscr{Q}$, may easily be seen to arise through Szegö's theorem (Sections 1.3.2.6.9.4, 1.3.4.2.1.10) from the extra logarithmic term in Stirling's formula

$$\log(q!) \approx q \log q - q + \tfrac{1}{2} \log(2\pi q)$$

(see, for instance, Reif, 1965) beyond the first two terms which serve to define entropy, since

$$\frac{1}{n} \log \det(2\pi \mathbf{Q}) \approx \int_{\mathbb{R}^3/\mathbb{Z}^3} \log 2\pi q^{\mathrm{ME}}(\mathbf{x}) \, \mathrm{d}^3\mathbf{x}.$$

The relative effect of this extra normalization factor depends on the ratio

$$\frac{n}{N} = \frac{\text{dimension of } \mathbf{F} \text{ over } \mathbb{R}}{\text{number of atoms}}.$$

The above relation between entropy maximization and the saddlepoint approximation is the basis of a Bayesian statistical approach to the phase problem (Bricogne, 1988) where the assumptions under which joint distributions of structure factors are sought incorporate many new ingredients (such as molecular boundaries, isomorphous substitutions, known fragments, noncrystallographic symmetries, multiple crystal forms) besides trial phase choices for basis reflections. The ME criterion intervenes in the construction of $q^{\mathrm{ME}}(\mathbf{x})$ under these assumptions, and the distribution $q^{\mathrm{ME}}(\mathbf{x})$ is a very useful computational intermediate in obtaining the approximate joint probability $\mathscr{P}^{\mathrm{SP}}(\mathbf{F}^*)$ and the associated conditional distributions and likelihood functions.

(*f*) *Role of the Fourier transformation*

The formal developments presented above make use of the following properties of the Fourier transformation:

(i) the convolution theorem, which turns the convolution of probability distributions into the multiplication of their characteristic functions;

(ii) the differentiation property, which confers moment-generating properties to characteristic functions;

(iii) the reciprocity theorem, which allows the retrieval of a probability distribution from its characteristic or moment-generating function;

(iv) the Paley–Wiener theorem, which allows the analytic continuation of characteristic functions associated to probability distributions with compact support, and thus gives rise to conjugate families of distributions;

(v) Bertaut's structure-factor algebra (a discrete symmetrized version of the convolution theorem), which allows the calculation of all necessary moments and cumulants when the dimension n is small;

(vi) Szegö's theorem, which provides an asymptotic approximation of the normalization factor when n is large.

This multi-faceted application seems an appropriate point at which to end this description of the Fourier transformation and of its use in crystallography.

Many aspects of the theory of discrete Fourier transform algorithms and of its extension to incorporate crystallographic symmetry have been the focus of a long-standing collaborative effort between Professor Louis Auslander, Professor Richard Tolimieri, their co-workers and the writer. I am most grateful to them for many years of mathematical stimulation and enjoyment, for introducing me to the 'big picture' of the discrete Fourier transform which they have elaborated over the past decade, and for letting me describe here some of their unpublished work. In particular, the crystallographic extensions of the Rader/Winograd algorithms presented in Section 1.3.4.3.4.3 were obtained by Richard Tolimieri, in a collaboration partially supported by NIH grant GM 32362 (to the writer). I am indebted to the Editor for many useful and constructive suggestions of possible improvements to the text, only a few of which I have been able to implement. I hope to incorporate many more of them in the future. I also wish to thank Dr D. Sayre for many useful comments on an early draft of the manuscript. This contribution was written during the tenure of a Visiting Fellowship at Trinity College, Cambridge, with partial financial support from Trinity College and the MRC Laboratory of Molecular Biology. I am most grateful to both institutions for providing ideal working conditions.

References

Agarwal, R. C. (1978). *A new least-squares refinement technique based on the fast Fourier transform algorithm. Acta Cryst.* A**34**, 791–809.

Agarwal, R. C. (1980). *The refinement of crystal structure by fast Fourier least-squares*. In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 18.01–18.13. Bangalore: The Indian Academy of Science.

Agarwal, R. C. (1981). *New results on fast Fourier least-squares refinement technique*. In *Refinement of Protein Structures*, compiled by P. A. Machin, J. W. Campbell & M. Elder (ref. DL/SCI/R16), pp. 24–28. Warrington: SERC Daresbury Laboratory.

Agarwal, R. C. & Cooley, J. W. (1986). *Fourier transform and convolution subroutines for the IBM 3090 Vector facility. IBM J. Res. Dev.* **30**, 145–162.

Agarwal, R. C. & Cooley, J. W. (1987). *Vectorized mixed radix discrete Fourier transform algorithms. Proc. IEEE*, **75**, 1283–1292.

Agarwal, R. C., Lifchitz, A. & Dodson, E. J. (1981). Appendix (pp. 36–38) to Dodson (1981).

Ahlfors, L. V. (1966). *Complex Analysis*. New York: McGraw-Hill.

Ahmed, F. R. & Barnes, W. H. (1958). *Generalized programmes for crystallographic computations. Acta Cryst*. **11**, 669–671.

Ahmed, F. R. & Cruickshank, D. W. J. (1953*a*). *A refinement of the crystal structure analysis of oxalic acid dihydrate. Acta Cryst.* **6**, 385–392.

Ahmed, F. R. & Cruickshank, D. W. J. (1953*b*). *Crystallographic calculations on the Manchester University electronic digital computer (Mark II). Acta Cryst.* **6**, 765–769.

Akhiezer, N. I. (1965). *The Classical Moment Problem*. Edinburgh, London: Oliver & Boyd.

Alston, N. A. & West, J. (1929). *The structure of topaz, [Al(F,OH)]$_2$SiO$_4$. Z. Kristallogr.* **69**, 149–167.

Apostol, T. M. (1976). *Introduction to Analytic Number Theory*. New York: Springer-Verlag.

Arnold, H. (2005). *Transformations in crystallography*. In *International Tables for Crystallography*, Vol. A. *Space-Group Symmetry*, edited by Th. Hahn, Chapter 5.1. Heidelberg: Springer.

Artin, E. (1944). *Galois Theory*. Notre Dame University Press.

Ascher, E. & Janner, A. (1965). *Algebraic aspects of crystallography. I. Space groups as extensions. Helv. Phys. Acta*, **38**, 551–572.

Ascher, E. & Janner, A. (1968). *Algebraic aspects of crystallography. II. Non-primitive translations in space groups. Commun. Math. Phys*. **11**, 138–167.

Ash, J. M. (1976). *Multiple trigonometric series*. In *Studies in Harmonic Analysis*, edited by J. M. Ash, pp. 76–96. *MAA studies in mathematics*, Vol. 13. The Mathematical Association of America.

Auslander, L. (1965). *An account of the theory of crystallographic groups. Proc. Am. Math. Soc*. **16**, 1230–1236.

Auslander, L., Feig, E. & Winograd, S. (1982). *New algorithms for evaluating the 3-dimensional discrete Fourier transform*. In *Computational Crystallography*, edited by D. Sayre, pp. 451–461. New York: Oxford University Press.

Auslander, L., Feig, E. & Winograd, S. (1983). *New algorithms for the multidimensional discrete Fourier transform. IEEE Trans. Acoust. Speech Signal Process*. **31**, 388–403.

Auslander, L., Feig, E. & Winograd, S. (1984). *Abelian semi-simple algebras and algorithms for the discrete Fourier transform. Adv. Appl. Math*. **5**, 31–55.

Auslander, L., Johnson, R. W. & Vulis, M. (1988). *Evaluating finite Fourier transforms that respect group symmetries. Acta Cryst.* A**44**, 467–478.

Auslander, L. & Shenefelt, M. (1987). *Fourier transforms that respect crystallographic symmetries. IBM J. Res. Dev*. **31**, 213–223.

Auslander, L. & Tolimieri, R. (1979). *Is computing with the finite Fourier transform pure or applied mathematics? Bull. Am. Math. Soc*. **1**, 847–897.

Auslander, L. & Tolimieri, R. (1985). *Ring structure and the Fourier transform. Math. Intelligencer*, **7**, 49–54.

Auslander, L., Tolimieri, R. & Winograd, S. (1982). *Hecke's theorem in quadratic reciprocity, finite nilpotent groups and the Cooley–Tukey algorithm. Adv. Math*. **43**, 122–172.

Ayoub, R. (1963). *An Introduction to the Analytic Theory of Numbers*. Providence: American Mathematical Society.

Baker, E. N. & Dodson, E. J. (1980). *Crystallographic refinement of the structure of actinidin at 1.7 Å resolution by fast Fourier least-squares methods. Acta Cryst.* A**36**, 559–572.

Bantz, D. & Zwick, M. (1974). *The use of symmetry with the fast Fourier algorithm. Acta Cryst.* A**30**, 257–260.

Barakat, R. (1974). *First-order statistics of combined random sinusoidal waves with applications to laser speckle patterns. Opt. Acta*, **21**, 903–921.

Barrett, A. N. & Zwick, M. (1971). *A method for the extension and refinement of crystallographic protein phases utilizing the fast Fourier transform. Acta Cryst.* A**27**, 6–11.

Beevers, C. A. (1952). *Fourier strips at a 3° interval. Acta Cryst.* **5**, 670–673.

Beevers, C. A. & Lipson, H. (1934). *A rapid method for the summation of a two-dimensional Fourier series. Philos. Mag*. **17**, 855–859.

Beevers, C. A. & Lipson, H. (1936). *A numerical method for two-dimensional Fourier synthesis. Nature (London)*, **137**, 825–826.

Beevers, C. A. & Lipson, H. (1952). *The use of Fourier strips for calculating structure factors. Acta Cryst.* **5**, 673–675.

Bellman, R. (1958). *Dynamic programming and stochastic control processes. Inf. Control*, **1**, 228–239.

Bennett, J. M. & Kendrew, J. C. (1952). *The computation of Fourier syntheses with a digital electronic calculating machine. Acta Cryst.* **5**, 109–116.

Berberian, S. K. (1962). *Measure and Integration*. New York: Macmillan. [Reprinted by Chelsea, New York, 1965.]

Bertaut, E. F. (1952). *L'énergie électrostatique de réseaux ioniques. J. Phys. Radium*, **13**, 499–505.

Bertaut, E. F. (1955*a*). *La méthode statistique en cristallographie. I. Acta Cryst*. **8**, 537–543.

Bertaut, E. F. (1955*b*). *La méthode statistique en cristallographie. II. Quelques applications. Acta Cryst.* **8**, 544–548.

Bertaut, E. F. (1955*c*). *Fonction de répartition: application à l'approache directe des structures. Acta Cryst.* **8**, 823–832.

Bertaut, E. F. (1956*a*). *Les groupes de translation non primitifs et la méthode statistique. Acta Cryst*. **9**, 322.

Bertaut, E. F. (1956*b*). *Tables de linéarisation des produits et puissances des facteurs de structure. Acta Cryst.* **9**, 322–323.

Bertaut, E. F. (1956*c*). *Algèbre des facteurs de structure. Acta Cryst.* **9**, 769–770.

Bertaut, E. F. (1959*a*). *IV. Algèbre des facteurs de structure. Acta Cryst.* **12**, 541–549.

Bertaut, E. F. (1959*b*). *V. Algèbre des facteurs de structure. Acta Cryst.* **12**, 570–574.

Bertaut, E. F. & Waser, J. (1957). *Structure factor algebra. II. Acta Cryst.* **10**, 606–607.

Bhattacharya, R. N. & Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*. New York: John Wiley.

Bieberbach, L. (1911). *Über die Bewegungsgruppen der Euklidischen Raume I. Math. Ann.* **70**, 297–336.

Bieberbach, L. (1912). *Über die Bewegungsgruppen der Euklidischen Raume II. Math. Ann.* **72**, 400–412.

Bienenstock, A. & Ewald, P. P. (1962). *Symmetry of Fourier space. Acta Cryst.* **15**, 1253–1261.

Blahut, R. E. (1985). *Fast Algorithms for Digital Signal Processing*. Reading: Addison-Wesley.

Bleistein, N. & Handelsman, R. A. (1986). *Asymptotic Expansions of Integrals*. New York: Dover Publications.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Ångström resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Bochner, S. (1932). *Vorlesungen über Fouriersche Integrale*. Leipzig: Akademische Verlagsgesellschaft.

Bochner, S. (1959). *Lectures on Fourier Integrals*. Translated from Bochner (1932) by M. Tenenbaum & H. Pollard. Princeton University Press.

Bode, W. & Schwager, P. (1975). *The refined crystal structure of bovine β-trypsin at 1.8Å resolution. II. Crystallographic refinement, calcium-binding site, benzamidine binding site and active site at pH 7.0. J. Mol. Biol.* **98**, 693–717.

Bondot, P. (1971). *Application de la transformée de Fourier performante aux calculs cristallographiques. Acta Cryst.* A**27**, 492–494.

Booth, A. D. (1945*a*). *Two new modifications of the Fourier method of X-ray structure analysis. Trans. Faraday Soc.* **41**, 434–438.

Booth, A. D. (1945*b*). *An expression for following the process of refinement in X-ray structure analysis using Fourier series. Philos. Mag.* **36**, 609–615.

Booth, A. D. (1945*c*). *Accuracy of atomic co-ordinates derived from Fourier synthesis. Nature (London)*, **156**, 51–52.

Booth, A. D. (1946*a*). *A differential Fourier method for refining atomic parameters in crystal structure analysis. Trans. Faraday Soc.* **42**, 444–448.

Booth, A. D. (1946*b*). *The simultaneous differential refinement of co-ordinates and phase angles in X-ray Fourier synthesis. Trans. Faraday Soc.* **42**, 617–619.

Booth, A. D. (1946*c*). *The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. Proc. R. Soc. London Ser. A*, **188**, 77–92.

Booth, A. D. (1947*a*). *The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. III. Proc. R. Soc. London Ser. A*, **190**, 482–489.

Booth, A. D. (1947*b*). *The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. IV. The two-dimensional projection of oxalic acid. Proc. R. Soc. London Ser. A*, **190**, 490–496.

Booth, A. D. (1947*c*). *A new refinement technique for X-ray structure analysis. J. Chem. Phys.* **15**, 415–416.

Booth, A. D. (1947*d*). *Application of the method of steepest descents to X-ray structure analysis. Nature (London)*, **160**, 196.

Booth, A. D. (1948*a*). *A new Fourier refinement technique. Nature (London)*, **161**, 765–766.

Booth, A. D. (1948*b*). *Fourier Technique in X-ray Organic Structure Analysis*. Cambridge University Press.

Booth, A. D. (1949). *The refinement of atomic parameters by the technique known in X-ray crystallography as 'the method of steepest descents'. Proc. R. Soc. London Ser. A*, **197**, 336–355.

Born, M. & Huang, K. (1954). *Dynamical Theory of Crystal Lattices*. Oxford University Press.

Bracewell, R. N. (1986). *The Fourier Transform and its Applications*, 2nd ed., revised. New York: McGraw-Hill.

Bragg, W. H. (1915). *X-rays and crystal structure*. (Bakerian Lecture.) *Philos. Trans. R. Soc. London Ser. A*, **215**, 253–274.

Bragg, W. L. (1929). *Determination of parameters in crystal structures by means of Fourier series. Proc. R. Soc. London Ser. A*, **123**, 537–559.

Bragg, W. L. (1975). *The Development of X-ray Analysis*, edited by D. C. Phillips & H. Lipson. London: Bell.

Bragg, W. L. & Lipson, H. (1936). *The employment of contoured graphs of structure-factor in crystal analysis. Z. Kristallogr.* **95**, 323–337.

Bragg, W. L. & West, J. (1929). *A technique for the X-ray examination of crystal structures with many parameters. Z. Kristallogr.* **69**, 118–148.

Bragg, W. L. & West, J. (1930). *A note on the representation of crystal structure by Fourier series. Philos. Mag.* **10**, 823–841.

Bremermann, H. (1965). *Distributions, Complex Variables, and Fourier Transforms*. Reading: Addison-Wesley.

Bricogne, G. (1974). *Geometric sources of redundancy in intensity data and their use for phase determination. Acta Cryst.* A**30**, 395–405.

Bricogne, G. (1976). *Methods and programs for direct-space exploitation of geometric redundancies. Acta Cryst.* A**32**, 832–847.

Bricogne, G. (1982). *Generalised density modification methods*. In *Computational Crystallography,* edited by D. Sayre, pp. 258–264. New York: Oxford University Press.

Bricogne, G. (1984). *Maximum entropy and the foundations of direct methods. Acta Cryst.* A**40**, 410–445.

Bricogne, G. (1988). *A Bayesian statistical theory of the phase problem. I. A multichannel maximum entropy formalism for constructing generalised joint probability distributions of structure factors. Acta Cryst.* A**44**, 517–545.

Bricogne, G. & Tolimieri, R. (1990). *Two-dimensional FFT algorithms on data admitting 90°-rotational symmetry*. In *Signal Processing Theory*, edited by L. Auslander, T. Kailath & S. Mitter, pp. 25–35. New York: Springer-Verlag.

Brigham, E. O. (1988). *The Fast Fourier Transform and its Applications*. Englewood Cliffs: Prentice-Hall.

Brill, R., Grimm, H., Hermann, C. & Peters, C. (1939). *Anwendung der röntgenographischen Fourieranalyse auf Fragen den chemischen Bindung. Ann. Phys. (Leipzig)*, **34**, 393–445.

Britten, P. L. & Collins, D. M. (1982). *Information theory as a basis for the maximum determinant. Acta Cryst.* A**38**, 129–132.

Brown, H. (1969). *An algorithm for the determination of space groups. Math. Comput.* **23**, 499–514.

Brown, H., Bülow, R., Neubüser, J., Wondratschek, H. & Zassenhaus, H. (1978). *Crystallographic Groups of Four-Dimensional Space*. New York: John Wiley.

Bruijn, N. G. de (1970). *Asymptotic Methods in Analysis*, 3rd ed. Amsterdam: North-Holland.

Brünger, A. T. (1988). *Crystallographic refinement by simulated annealing*. In *Crystallographic Computing 4: Techniques and New Technologies*, edited by N. W. Isaacs & M. R. Taylor, pp. 126–140. New York: Oxford University Press.

Brünger, A. T. (1989). *A memory-efficient fast Fourier transformation algorithm for crystallographic refinement on supercomputers. Acta Cryst.* A**45**, 42–50.

Brünger, A. T., Karplus, M. & Petsko, G. A. (1989). *Crystallographic refinement by simulated annealing: application to crambin. Acta Cryst.* A**45**, 50–61.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics. Science*, **235**, 458–460.

Bryan, R. K., Bansal, M., Folkhard, W., Nave, C. & Marvin, D. A. (1983). *Maximum-entropy calculation of the electron density at 4 Å resolution of Pf1 filamentous bacteriophage. Proc. Natl Acad. Sci. USA*, **80**, 4728–4731.

Bryan, R. K. & Skilling, J. (1980). *Deconvolution by maximum entropy, as illustrated by application to the jet of M87. Mon. Not. R. Astron. Soc.* **191**, 69–79.

Burch, S. F., Gull, S. F. & Skilling, J. (1983). *Image restoration by a powerful maximum entropy method. Comput. Vision Graphics Image Process.* **23**, 113–128.

Burnett, R. M. & Nordman, C. E. (1974). *Optimization of the calculation of structure factors for large molecules. J. Appl. Cryst.* **7**, 625–627.

Burnside, W. (1911). *Theory of Groups of Finite Order*, 2nd ed. Cambridge University Press.

Burrus, C. S. & Eschenbacher, P. W. (1981). *An in-place, in-order prime factor FFT algorithm. IEEE Trans. Acoust. Speech Signal Process.* **29**, 806–817.

Busing, W. R. & Levy, H. A. (1961). *Least squares refinement programs for the IBM 704*. In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 146–149. Oxford: Pergamon Press.

Byerly, W. E. (1893). *An Elementary Treatise on Fourier's Series and Spherical, Cylindrical and Ellipsoidal Harmonics*. Boston: Ginn & Co. [Reprinted by Dover Publications, New York, 1959.]

Campbell, G. A. & Foster, R. M. (1948). *Fourier Integrals for Practical Applications*. Princeton: Van Nostrand.

Carathéodory, C. (1911). *Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Functionen. Rend. Circ. Mat. Palermo*, **32**, 193–217.

Carslaw, H. S. (1930). *An Introduction to the Theory of Fourier's Series and Integrals*. London: Macmillan. [Reprinted by Dover Publications, New York, 1950.]

Carslaw, H. S. & Jaeger, J. C. (1948). *Operational Methods in Applied Mathematics*. Oxford University Press.

Cartan, H. (1961). *Théorie des Fonctions Analytiques*. Paris: Hermann.

Challifour, J. L. (1972). *Generalized Functions and Fourier Analysis*. Reading: Benjamin.

Champeney, D. C. (1973). *Fourier Transforms and their Physical Applications*. New York: Academic Press.

Churchill, R. V. (1958). *Operational Mathematics*, 2nd ed. New York: McGraw-Hill.

Cochran, W. (1948*a*). *A critical examination of the Beevers–Lipson method of Fourier series summation. Acta Cryst.* **1**, 54–56.

Cochran, W. (1948*b*). *The Fourier method of crystal structure analysis. Nature (London)*, **161**, 765.

Cochran, W. (1948*c*). *The Fourier method of crystal-structure analysis. Acta Cryst.* **1**, 138–142.

Cochran, W. (1948*d*). *X-ray analysis and the method of steepest descents. Acta Cryst.* **1**, 273.

Cochran, W. (1951*a*). *Some properties of the $F_o - F_c$-synthesis. Acta Cryst.* **4**, 408–411.

Cochran, W. (1951*b*). *The structures of pyrimidines and purines. V. The electron distribution in adenine hydrochloride. Acta Cryst.* **4**, 81–92.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *The structure of synthetic polypeptides. I. The transform of atoms on a helix. Acta Cryst.* **5**, 581–586.

Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., Maling, G. C., Nelson, D. E., Rader, C. M. & Welch, P. D. (1967). *What is the fast Fourier transform? IEEE Trans. Audio*, **15**, 45–55.

Collins, D. M. (1975). *Efficiency in Fourier phase refinement for protein crystal structures. Acta Cryst.* A**31**, 388–389.

Collins, D. M., Brice, M. D., Lacour, T. F. M. & Legg, M. J. (1976). *Fourier phase refinement and extension by modification of electron-density maps*. In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 330–335. Copenhagen: Munksgaard.

Colman, P. M. (1974). *Non-crystallographic symmetry and the sampling theorem. Z. Kristallogr.* **140**, 344–349.

Cooley, J. W. & Tukey, J. W. (1965). *An algorithm for the machine calculation of complex Fourier series. Math. Comput.* **19**, 297–301.

Cox, E. G. & Cruickshank, D. W. J. (1948). *The accuracy of electron-density maps in X-ray structure analysis. Acta Cryst.* **1**, 92–93.
Cox, E. G., Gross, L. & Jeffrey, G. A. (1947). *A Hollerith punched-card method for the evaluation of electron density in crystal structure analysis. Proc. Leeds Philos. Soc.* **5**, 1–13.
Cox, E. G., Gross, L. & Jeffrey, G. A. (1949). *A Hollerith technique for computing three-dimensional differential Fourier syntheses in X-ray crystal structure analysis. Acta Cryst.* **2**, 351–355.
Cox, E. G. & Jeffrey, G. A. (1949). *The use of Hollerith computing equipment in crystal structure analysis. Acta Cryst.* **2**, 341–343.
Coxeter, H. S. M. & Moser, W. O. J. (1972). *Generators and Relations for Discrete Groups*, 3rd ed. Berlin: Springer-Verlag.
Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
Crowther, R. A. (1967). *A linear analysis of the non-crystallographic symmetry problem. Acta Cryst.* **22**, 758–764.
Crowther, R. A. (1969). *The use of non-crystallographic symmetry for phase determination. Acta Cryst.* B**25**, 2571–2580.
Crowther, R. A. (1972). *The fast rotation function*. In *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.
Cruickshank, D. W. J. (1949*a*). *The accuracy of electron-density maps in X-ray analysis with special reference to dibenzyl. Acta Cryst.* **2**, 65–82.
Cruickshank, D. W. J. (1949*b*). *The accuracy of atomic co-ordinates derived by least-squares or Fourier methods. Acta Cryst.* **2**, 154–157.
Cruickshank, D. W. J. (1950). *The convergence of the least-squares and Fourier refinement methods. Acta Cryst.* **3**, 10–13.
Cruickshank, D. W. J. (1952). *On the relations between Fourier and least-squares methods of structure determination. Acta Cryst.* **5**, 511–518.
Cruickshank, D. W. J. (1956). *The determination of the anisotropic thermal motion of atoms in crystals. Acta Cryst.* **9**, 747–753.
Cruickshank, D. W. J. (1965*a*). *Errors in Fourier series*. In *Computing Methods in Crystallography*, edited by J. S. Rollett, pp. 107–111. Oxford: Pergamon Press.
Cruickshank, D. W. J. (1965*b*). *Errors in least-squares methods*. In *Computing Methods in Crystallography*, edited by J. S. Rollett, pp. 112–116. Oxford: Pergamon Press.
Cruickshank, D. W. J. (1970). *Least-squares refinement of atomic parameters*. In *Crystallographic Computing*, edited by F. R. Ahmed, pp. 187–197. Copenhagen: Munksgaard.
Cruickshank, D. W. J., Pilling, D. E., Bujosa, A., Lovell, F. M. & Truter, M. R. (1961). *Crystallographic calculations on the Ferranti Pegasus and mark I computers*. In *Computing Methods and the Phase Problem in X-Ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 32–78. Oxford: Pergamon Press.
Cruickshank, D. W. J. & Rollett, J. S. (1953). *Electron-density errors at special positions. Acta Cryst.* **6**, 705–707.
Curtis, C. W. & Reiner, I. (1962). *Representation Theory of Finite Groups and Associative Algebras*. New York: Wiley–Interscience.
Daniels, H. E. (1954). *Saddlepoint approximation in statistics. Ann. Math. Stat.* **25**, 631–650.
Deisenhofer, J. & Steigemann, W. (1975). *Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Cryst.* B**31**, 238–250.
Diamond, R. (1971). *A real-space refinement procedure for proteins. Acta Cryst.* A**27**, 436–452.
Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961*a*). *The phase problem and isomorphous replacement methods in protein structures*. In *Computing Methods and the Phase Problem in X-Ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 236–251. Oxford: Pergamon Press.
Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961*b*). *The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. Acta Cryst.* **14**, 1188–1195.
Dietrich, H. (1972). *A reconsideration of Fourier methods for the refinement of crystal structures. Acta Cryst.* B**28**, 2807–2814.
Dieudonné, J. (1969). *Foundations of Modern Analysis*. New York, London: Academic Press.
Dieudonné, J. (1970). *Treatise on Analysis*, Vol. II. New York, London: Academic Press.
Dirac, P. A. M. (1958). *The Principles of Quantum Mechanics*, 4th ed. Oxford: Clarendon Press.
Dodson, E. J. (1981). *Block diagonal least squares refinement using fast Fourier techniques*. In *Refinement of Protein Structures*, compiled by P. A. Machin J. W. Campbell & M. Elder (ref. DL/SCI/R16), pp. 29–39. Warrington: SERC Daresbury Laboratory.
Donohue, J. & Schomaker, V. (1949). *The use of punched cards in molecular structure determinations*. III. *Structure factor calculations of X-ray crystallography. Acta Cryst.* **2**, 344–347.
Duane, W. (1925). *The calculation of the X-ray diffracting power at points in a crystal. Proc. Natl Acad. Sci. USA*, **11**, 489–493.
Dym, H. & McKean, H. P. (1972). *Fourier Series and Integrals*. New York, London: Academic Press.
Eklundh, J. O. (1972). *A fast computer method for matrix transposing. IEEE Trans*. C-**21**, 801–803.
Engel, P. (1986). *Geometric Crystallography*. Dordrecht: Kluwer Academic Publishers.
Erdélyi, A. (1954). *Tables of Integral Transforms*, Vol. I. New York: McGraw-Hill.
Erdélyi, A. (1962). *Operational Calculus and Generalized Functions*. New York: Holt, Rinehart & Winston.
Ewald, P. P. (1921). *Die Berechnung optischer und electrostatischer Gitterpotentiale. Ann. Phys. Leipzig*, **64**, 253–287.
Ewald, P. P. (1940). *X-ray diffraction by finite and imperfect crystal lattices. Proc. Phys. Soc. London*, **52**, 167–174.
Ewald, P. P. (1962). *Fifty Years of X-ray Diffraction*. Dordrecht: Kluwer Academic Publishers.
Farkas, D. R. (1981). *Crystallographic groups and their mathematics. Rocky Mountain J. Math.* **11**, 511–551.
Fornberg, A. (1981). *A vector implementation of the fast Fourier transform algorithm. Math. Comput.* **36**, 189–191.
Forsyth, J. B. & Wells, M. (1959). *On an analytical approximation to the atomic scattering factor. Acta Cryst.* **12**, 412–415.
Fowler, R. H. (1936). *Statistical Mechanics*, 2nd ed. Cambridge University Press.
Fowweather, F. (1955). *The use of general programmes for crystallographic calculations on the Manchester University electronic digital computer (Mark II). Acta Cryst.* **8**, 633–637.
Freer, S. T., Alden, R. A., Levens, S. A. & Kraut, J. (1976). *Refinement of five protein structures by constrained $F_o - F_c$ Fourier methods*. In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 317–321. Copenhagen: Munksgaard.
Friedel, G. (1913). *Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen. C. R. Acad. Sci. Paris*, **157**, 1533–1536.
Friedlander, F. G. (1982). *Introduction to the Theory of Distributions*. Cambridge University Press.
Friedlander, P. H., Love, W. & Sayre, D. (1955). *Least-squares refinement at high speed. Acta Cryst.* **8**, 732.
Friedman, A. (1970). *Foundations of Modern Analysis*. New York: Holt, Rinehart & Winston. [Reprinted by Dover, New York, 1982.]
Frobenius, G. (1911). *Über die unzerlegbaren diskreten Bewegungsgruppen. Sitzungsber. Preuss. Akad. Wiss. Berlin*, **29**, 654–665.
Gassmann, J. (1976). *Improvement and extension of approximate phase sets in structure determination*. In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 144–154. Copenhagen: Munksgaard.
Gassmann, J. & Zechmeister, K. (1972). *Limits of phase expansion in direct methods. Acta Cryst.* A**28**, 270–280.
Gel'fand, I. M. & Shilov, G. E. (1964). *Generalized Functions*, Vol. I. New York, London: Academic Press.
Gentleman, W. M. & Sande, G. (1966). *Fast Fourier transforms – for fun and profit*. In *AFIPS Proc. 1966 Fall Joint Computer Conference*, pp. 563–578. Washington: Spartan Books.
Gillis, J. (1948*a*). *Structure factor relations and phase determination. Acta Cryst.* **1**, 76–80.
Gillis, J. (1948*b*). *The application of the Harker–Kasper method of phase determination. Acta Cryst.* **1**, 174–179.
Goedkoop, J. A. (1950). *Remarks on the theory of phase-limiting inequalities and equalities. Acta Cryst.* **3**, 374–378.
Goldstine, H. H. (1977). *A History of Numerical Analysis From the 16th Through the 19th Century*. New York: Springer-Verlag.
Good, I. J. (1958). *The interaction algorithm and practical Fourier analysis. J. R. Stat. Soc.* B**20**, 361–372.
Good, I. J. (1960). *Addendum* [to Good (1958)]. *J. R. Stat. Soc.* B**22**, 372–375.
Good, I. J. (1971). *The relationship between two fast Fourier transforms. IEEE Trans*. C-**20**, 310–317.

Greenhalgh, D. M. S. & Jeffrey, G. A. (1950). *A new punched card method of Fourier synthesis. Acta Cryst.* **3**, 311–312.

Grems, M. D. & Kasper, J. S. (1949). *An improved punched-card method for crystal structure-factor calculations. Acta Cryst.* **2**, 347–351.

Grenander, U. (1952). *On Toeplitz forms and stationary processes. Ark. Math.* **1**, 555–571.

Grenander, U. & Szegö, G. (1958). *Toeplitz Forms and Their Applications*. Berkeley: University of California Press.

Guessoum, A. & Mersereau, R. M. (1986). *Fast algorithms for the multidimensional discrete Fourier transform. IEEE Trans. Acoust. Speech Signal Process.* **34**, 937–943.

Hadamard, J. (1932). *Le Problème de Cauchy et les Equations aux Dérivées Partielles Linéaires Hyperboliques*. Paris: Hermann.

Hadamard, J. (1952). *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New York: Dover Publications.

Hall, M. (1959). *The Theory of Groups*. New York: Macmillan.

Hardy, G. H. (1933). *A theorem concerning Fourier transforms. J. London Math. Soc.* **8**, 227–231.

Harker, D. (1936). *The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. J. Chem. Phys.* **4**, 381–390.

Harker, D. & Kasper, J. S. (1948). *Phases of Fourier coefficients directly from crystal diffraction data. Acta Cryst.* **1**, 70–75.

Harris, D. B., McClellan, J. H., Chan, D. S. K. & Schuessler, H. W. (1977). *Vector radix fast Fourier transform. Rec. 1977 IEEE Internat. Conf. Acoust. Speech Signal Process*. pp. 548–551.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Tomato bushy stunt virus at 2.9 Ångström resolution. Nature (London)*, **276**, 368–373.

Hartman, P. & Wintner, A. (1950). *On the spectra of Toeplitz's matrices. Am. J. Math.* **72**, 359–366.

Hartman, P. & Wintner, A. (1954). *The spectra of Toeplitz's matrices. Am. J. Math.* **76**, 867–882.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*. ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.

Havighurst, R. J. (1925*a*). *The distribution of diffracting power in sodium chloride. Proc. Natl Acad. Sci. USA*, **11**, 502–507.

Havighurst, R. J. (1925*b*). *The distribution of diffracting power in certain crystals. Proc. Natl Acad. Sci. USA*, **11**, 507–512.

Heideman, M. T., Johnson, D. H. & Burrus, C. S. (1984). *Gauss and the history of the fast Fourier transform. IEEE Acoust. Speech Signal Process. Mag.* October 1984.

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of stereochemical information into crystallographic refinement*. In *Computing in Crystallography,* edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.26. Bangalore: The Indian Academy of Science.

Herglotz, G. (1911). *Über Potenzreihen mit positiven, reellen Teil im Einheitskreis. Ber. Sächs. Ges. Wiss. Leipzig*, **63**, 501–511.

Hirschman, I. I. Jr & Hughes, D. E. (1977). *Extreme Eigenvalues of Toeplitz Operators. Lecture Notes in Mathematics*, Vol. 618. Berlin: Springer-Verlag.

Hodgson, M. L., Clews, C. J. B. & Cochran, W. (1949). *A punched card modification of the Beevers–Lipson method of Fourier synthesis. Acta Cryst.* **2**, 113–116.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Hoppe, W., Gassmann, J. & Zechmeister, K. (1970). *Some automatic procedures for the solution of crystal structures with direct methods and phase correction.* In *Crystallographic Computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 26–36. Copenhagen: Munksgaard.

Hörmander, L. (1963). *Linear Partial Differential Operators*. Berlin: Springer-Verlag.

Hörmander, L. (1973). *An Introduction to Complex Analysis in Several Variables*, 2nd ed. Amsterdam: North-Holland.

Huber, R., Kulka, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). *Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. J. Mol. Biol.* **89**, 73–101.

Hughes, E. W. (1941). *The crystal structure of melamine. J. Am. Chem. Soc.* **63**, 1737–1752.

Immirzi, A. (1973). *A general Fourier program for X-ray crystal-structure analysis which utilizes the Cooley–Tukey algorithm. J. Appl. Cryst.* **6**, 246–249.

Immirzi, A. (1976). *Fast Fourier transform in crystallography*. In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 399–412. Copenhagen: Munksgaard.

Isaacs, N. W. (1982*a*). *The refinement of macromolecules*. In *Computational Crystallography*, edited by D. Sayre, pp. 381–397. New York: Oxford University Press.

Isaacs, N. W. (1982*b*). *Refinement techniques: use of the FFT*. In *Computational Crystallography*, edited by D. Sayre, pp. 398–408. New York: Oxford University Press.

Isaacs, N. W. (1984). *Refinement using the fast-Fourier transform least squares algorithm*. In *Methods and Applications in Crystallographic Computing*, edited by S. R. Hall & T. Ashida, pp. 193–205. New York: Oxford University Press.

Isaacs, N. W. & Agarwal, R. C. (1978). *Experience with fast Fourier least squares in the refinement of the crystal structure of rhombohedral 2-zinc insulin at 1.5 Å resolution. Acta Cryst*. A**34**, 782–791.

Jack, A. (1973). *Direct determination of X-ray phases for tobacco mosaic virus protein using non-crystallographic symmetry. Acta Cryst*. A**29**, 545–554.

Jack, A. & Levitt, M. (1978). *Refinement of large structures by simultaneous minimization of energy and R factor. Acta Cryst*. A**34**, 931–935.

James, R. W. (1948*a*). *The Optical Principles of the Diffraction of X-rays*. London: Bell.

James, R. W. (1948*b*). *False detail in three-dimensional Fourier representations of crystal structures. Acta Cryst.* **1**, 132–134.

Janssen, T. (1973). *Crystallographic Groups*. Amsterdam: North-Holland.

Jaynes, E. T. (1957). *Information theory and statistical mechanics. Phys. Rev.* **106**, 620–630.

Jaynes, E. T. (1968). *Prior probabilities. IEEE Trans. SSC*, **4**, 227–241.

Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*. Dordrecht: Kluwer Academic Publishers.

Johnson, H. W. & Burrus, C. S. (1983). *The design of optimal DFT algorithms using dynamic programming. IEEE Trans. Acoust. Speech Signal Process*. **31**, 378–387.

Jürgensen, H. (1970). *Calculation with the elements of a finite group given by generators and defining relations*. In *Computational Problems in Abstract Algebra*, edited by J. Leech, pp. 47–57. Oxford: Pergamon Press.

Kac, M. (1954). *Toeplitz matrices, translation kernels, and a related problem in probability theory. Duke Math. J.* **21**, 501–509.

Kac, M., Murdock, W. L. & Szegö, G. (1953). *On the eigenvalues of certain Hermitian forms. J. Rat. Mech. Anal.* **2**, 767–800.

Karle, J. & Hauptman, H. (1950). *The phases and magnitudes of the structure factors. Acta Cryst.* **3**, 181–187.

Katznelson, Y. (1968). *An Introduction to Harmonic Analysis*. New York: John Wiley.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. New York: Dover Publications.

Kitz, N. & Marchington, B. (1953). *A method of Fourier synthesis using a standard Hollerith senior rolling total tabulator. Acta Cryst.* **6**, 325–326.

Klug, A. (1958). *Joint probability distributions of structure factors and the phase problem. Acta Cryst.* **11**, 515–543.

Klug, A., Crick, F. H. C. & Wyckoff, H. W. (1958). *Diffraction by helical structures. Acta Cryst.* **11**, 199–213.

Kluyver, J. C. (1906). *A local probability problem. K. Ned. Akad. Wet. Proc.* **8**, 341–350.

Kolba, D. P. & Parks, T. W. (1977). *A prime factor FFT algorithm using high-speed convolution. IEEE Trans. Acoust. Speech Signal Process*. **25**, 281–294.

Konnert, J. H. (1976). *A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. Acta Cryst.* A**32**, 614–617.

Konnert, J. H. & Hendrickson, W. A. (1980). *A restrained-parameter thermal-factor refinement procedure. Acta Cryst*. A**36**, 344–350.

Korn, D. G. & Lambiotte, J. J. Jr (1979). *Computing the fast Fourier transform on a vector computer. Math. Comput.* **33**, 977–992.

Kuriyan, J., Brünger, A. T., Karplus, M. & Hendrickson, W. A. (1989). *X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. Acta Cryst*. A**45**, 396–409.

Lanczos, C. (1966). *Discourse on Fourier Series*. Edinburgh: Oliver & Boyd.

Landau, H. J. & Pollack, H. O. (1961). *Prolate spheroidal wave functions, Fourier analysis and uncertainty (2). Bell Syst. Tech. J.* **40**, 65–84.

Landau, H. J. & Pollack, H. O. (1962). *Prolate spheroidal wave functions, Fourier analysis and uncertainty (3): the dimension of the space of essentially time- and band-limited signals. Bell Syst. Tech. J.* **41**, 1295–1336.

Lang, S. (1965). *Algebra*. Reading: Addison-Wesley.

Larmor, J. (1934). *The Fourier discontinuities: a chapter in historical integral calculus. Philos. Mag.* **17**, 668–678.

Laue, M. von (1936). *Die aüssere Form der Kristalle in ihrem Einfluss auf die Interferenzerscheinungen an Raumgittern. Ann. Phys. (Leipzig)*, **26**, 55–68.

Lavoine, J. (1963). *Transformation de Fourier des Pseudo-fonctions, avec Tables de Nouvelles Transformées*. Paris: Editions du CNRS.

Ledermann, W. (1987). *Introduction to Group Characters*, 2nd ed. Cambridge University Press.

Leslie, A. G. W. (1987). *A reciprocal-space method for calculating a molecular envelope using the algorithm of B. C. Wang. Acta Cryst.* A**43**, 134–136.

Lighthill, M. J. (1958). *Introduction to Fourier Analysis and Generalized Functions*. Cambridge University Press.

Linnik, I. Ju. (1975). *A multidimensional analogue of a limit theorem of G. Szegö. Math. USSR Izv*. **9**, 1323–1332.

Lipson, H. & Beevers, C. A. (1936). *An improved numerical method of two-dimensional Fourier synthesis for crystals. Proc. Phys. Soc. London*, **48**, 772–780.

Lipson, H. & Cochran, W. (1953). *The Determination of Crystal Structures*. London: Bell.

Lipson, H. & Cochran, W. (1968). *The Determination of Crystal Structures*. Revised and enlarged edition. London: G. Bell & Sons.

Lipson, H. & Taylor, C. A. (1951). *Optical methods in X-ray analysis. II. Fourier transforms and crystal-structure determination. Acta Cryst.* **4**, 458–462.

Lipson, H. & Taylor, C. A. (1958). *Fourier Transforms and X-ray Diffraction*. London: Bell.

Livesey, A. K. & Skilling, J. (1985). *Maximum entropy theory. Acta Cryst*. A**41**, 113–122.

Lonsdale, K. (1936). *Simplified Structure Factor and Electron Density Formulae for the 230 Space Groups of Mathematical Crystallography*. London: Bell.

Lunin, V. Yu. (1985). *Use of the fast differentiation algorithm for phase refinement in protein crystallography. Acta Cryst*. A**41**, 551–556.

McClellan, J. H. & Rader, C. M. (1979). *Number Theory in Digital Signal Processing*. Englewood Cliffs: Prentice Hall.

MacGillavry, C. H. (1950). *On the derivation of Harker–Kasper inequalities. Acta Cryst*. **3**, 214–217.

MacLane, S. (1963). *Homology*. Berlin: Springer-Verlag.

Magnus, W., Karrass, A. & Solitar, D. (1976). *Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations*, 2nd revised ed. New York: Dover Publications.

Magnus, W., Oberhettinger, F. & Soni, R. P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics*. Berlin: Springer-Verlag.

Main, P. & Rossmann, M. G. (1966). *Relationships among structure factors due to identical molecules in different crystallographic environments. Acta Cryst*. **21**, 67–72.

Main, P. & Woolfson, M. M. (1963). *Direct determination of phases by the use of linear equations between structure factors. Acta Cryst*. **16**, 1046–1051.

Mayer, S. W. & Trueblood, K. N. (1953). *Three-dimensional Fourier summations on a high-speed digital computer. Acta Cryst*. **6**, 427.

Mersereau, R. & Speake, T. C. (1981). *A unified treatment of Cooley–Tukey algorithms for the evaluation of the multidimensional DFT. IEEE Trans. Acoust. Speech Signal Process*. **29**, 1011–1018.

Mersereau, R. M. (1979). *The processing of hexagonally sampled two-dimensional signals. Proc. IEEE*, **67**, 930–949.

Montroll, E. W., Potts, R. B. & Ward, J. C. (1963). *Correlations and spontaneous magnetization of the two-dimensional Ising model. J. Math. Phys.* **4**, 308–322.

Moore, D. H. (1971). *Heaviside Operational Calculus. An Elementary Foundation*. New York: American Elsevier.

Morris, R. L. (1978). *A comparative study of time efficient FFT and WFTA programs for general purpose computers. IEEE Trans. Acoust. Speech Signal Process*. **26**, 141–150.

Narayan, R. & Nityananda, R. (1982). *The maximum determinant method and the maximum entropy method. Acta Cryst*. A**38**, 122–128.

Natterer, F. (1986). *The Mathematics of Computerized Tomography*. New York: John Wiley.

Navaza, J. (1985). *On the maximum-entropy estimate of the electron density function. Acta Cryst*. A**41**, 232–244.

Nawab, H. & McClellan, J. H. (1979). *Bounds on the minimum number of data transfers in WFTA and FFT programs. IEEE Trans. Acoust. Speech Signal Process*. **27**, 393–398.

Niggli, A. (1961). *Small-scale computers in X-ray crystallography*. In *Computing Methods and the Phase Problem*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 12–20. Oxford: Pergamon Press.

Nussbaumer, H. J. (1981). *Fast Fourier Transform and Convolution Algorithms*. Berlin: Springer-Verlag.

Nussbaumer, H. J. & Quandalle, P. (1979). *Fast computation of discrete Fourier transforms using polynomial transforms. IEEE Trans. Acoust. Speech Signal Process*. **27**, 169–181.

Onsager, L. (1944). *Crystal statistics. I. Two-dimensional model with an order–disorder transition. Phys. Rev*. **65**, 117–149.

Paley, R. E. A. C. & Wiener, N. (1934). *Fourier Transforms in the Complex Domain*. Providence: American Mathematical Society.

Patterson, A. L. (1934). *A Fourier series method for the determination of the components of interatomic distances in crystals. Phys. Rev*. **46**, 372–376.

Patterson, A. L. (1935*a*). *A direct method for the determination of the components of interatomic distances in crystals. Z. Kristallogr*. **90**, 517–542.

Patterson, A. L. (1935*b*). *Tabulated data for the seventeen plane groups. Z. Kristallogr.* **90**, 543–554.

Patterson, A. L. (1959). In *International Tables for X-ray Crystallography*, Vol. II, pp. 9–10. Erratum, January 1962. Birmingham: Kynoch Press.

Pearson, K. (1905). *The problem of the random walk. Nature (London)*, **72**, 294, 342.

Pease, M. C. (1968). *An adaptation of the fast Fourier transform for parallel processing. J. Assoc. Comput. Mach.* **15**, 252–264.

Pepinsky, R. (1947). *An electronic computer for X-ray crystal structure analyses. J. Appl. Phys.* **18**, 601–604.

Pepinsky, R. (1952). *The use of positive kernels in Fourier syntheses of crystal structures*. In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, pp. 319–338. State College: Penn. State University.

Pepinsky, R., van den Hende, J. & Vand, V. (1961). *X-RAC and digital computing methods*. In *Computing Methods and the Phase Problem*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 154–160. Oxford: Pergamon Press.

Pepinsky, R. & Sayre, D. (1948). *Quantitative electron-density contour delineation in the electronic Fourier synthesizer for crystal structure analysis. Nature (London)*, **162**, 22–23.

Petrov, V. V. (1975). *Sums of Independent Random Variables*. Berlin: Springer-Verlag.

Pollack, H. O. & Slepian, D. (1961). *Prolate spheroidal wave functions, Fourier analysis and uncertainty (1). Bell Syst. Tech. J.* **40**, 43–64.

Qurashi, M. M. (1949). *Optimal conditions for convergence of steepest descents as applied to structure determination. Acta Cryst*. **2**, 404–409.

Qurashi, M. M. (1953). *An analysis of the efficiency of convergence of different methods of structure determination. I. The methods of least squares and steepest descents: centrosymmetric case. Acta Cryst*. **6**, 577–588.

Qurashi, M. M. & Vand, V. (1953). *Weighting of the least-squares and steepest-descents methods in the initial stages of the crystal-structure determination. Acta Cryst*. **6**, 341–349.

Rader, C. M. (1968). *Discrete Fourier transforms when the number of data samples is prime. Proc. IEEE*, **56**, 1107–1108.

Rayleigh (J. W. Strutt), Lord (1880). *On the resultant of a large number of vibrations of the same pitch and arbitrary phase. Philos. Mag.* **10**, 73–78.

Rayleigh (J. W. Strutt), Lord (1899). *On James Bernoulli's theorem in probabilities. Philos. Mag.* **47**, 246–251.

Rayleigh (J. W. Strutt), Lord (1905). *The problem of the random walk. Nature (London)*, **72**, 318.

Rayleigh (J. W. Strutt), Lord (1918). *On the light emitted from a random distribution of luminous sources. Philos. Mag.* **36**, 429–449.

Rayleigh (J. W. Strutt), Lord (1919). *On the problem of random flights in one, two or three dimensions. Philos. Mag.* **37**, 321–347.

Reif, F. (1965). *Fundamentals of Statistical and Thermal Physics*, Appendix A.6. New York: McGraw-Hill.

Reijen, L. L. van (1942). *Diffraction effects in Fourier syntheses and their elimination in X-ray structure investigations. Physica*, **9**, 461–480.

Rice, S. O. (1944, 1945). *Mathematical analysis of random noise. Bell Syst. Tech. J.* **23**, 283–332 (parts I and II); **24**, 46–156 (parts III and IV). [Reprinted in *Selected Papers on Noise and Stochastic Processes* (1954), edited by N. Wax, pp. 133–294. New York: Dover Publications.]

Riesz, M. (1938). *L'intégrale de Riemann–Liouville et le problème de Cauchy pour l'équation des ondes. Bull. Soc. Math. Fr.* **66**, 153–170.

Riesz, M. (1949). *L'intégrale de Riemann–Liouville et le problème de Cauchy. Acta Math.* **81**, 1–223.

Rivard, G. E. (1977). *Direct fast Fourier transform of bivariate functions. IEEE Trans. Acoust. Speech Signal Process.* **25**, 250–252.

Robertson, J. M. (1932). *A simple harmonic continuous calculating machine. Philos. Mag.* **13**, 413–419.

Robertson, J. M. (1935). *An X-ray study of the structure of phthalocyanines. Part I. The metal-free, nickel, copper, and platinum compounds. J. Chem. Soc.* pp. 615–621.

Robertson, J. M. (1936*a*). *Numerical and mechanical methods in double Fourier synthesis. Philos. Mag.* **21**, 176–187.

Robertson, J. M. (1936*b*). *Calculation of structure factors and summation of Fourier series in crystal analysis: non-centrosymmetrical projections. Nature (London)*, **138**, 683–684.

Robertson, J. M. (1936*c*). *An X-ray study of the phthalocyanines. Part II. Quantitative structure determination of the metal-free compound. J. Chem. Soc.* pp. 1195–1209.

Robertson, J. M. (1937). *X-ray analysis and application of Fourier series methods to molecular structures. Phys. Soc. Rep. Prog. Phys.* **4**, 332–367.

Robertson, J. M. (1954). *A fast digital computer for Fourier operations. Acta Cryst.* **7**, 817–822.

Robertson, J. M. (1955). *Some properties of Fourier strips with applications to the digital computer. Acta Cryst.* **8**, 286–288.

Robertson, J. M. (1961). *A digital mechanical computer for Fourier operations*. In *Computing Methods and the Phase Problem,* edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 21–24. Oxford: Pergamon Press.

Rollett, J. S. (1965). *Structure factor routines*. In *Computing Methods in Crystallography*, edited by J. S. Rollett, pp. 38–46. Oxford: Pergamon Press.

Rollett, J. S. (1970). *Least-squares procedures in crystal structure analysis*. In *Crystallographic Computing*, edited by F. R. Ahmed, pp. 167–181. Copenhagen: Munksgaard.

Rollett, J. S. & Davies, D. R. (1955). *The calculation of structure factors for centrosymmetric monoclinic systems with anisotropic atomic vibration. Acta Cryst.* **8**, 125–128.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst.* **15**, 24–31.

Rossmann, M. G. & Blow, D. M. (1963). *Determination of phases by the conditions of non-crystallographic symmetry. Acta Cryst.* **16**, 39–45.

Sayre, D. (1951). *The calculation of structure factors by Fourier summation. Acta Cryst.* **4**, 362–367.

Sayre, D. (1952*a*). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Sayre, D. (1952*b*). *Some implications of a theorem due to Shannon. Acta Cryst.* **5**, 843.

Sayre, D. (1952*c*). *The Fourier transform in X-ray crystal analysis*. In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, pp. 361–390. State College: Penn. State University.

Sayre, D. (1972). *On least-squares refinement of the phases of crystallographic structure factors. Acta Cryst.* A**28**, 210–212.

Sayre, D. (1974). *Least-squares phase refinement. II. High-resolution phasing of a small protein. Acta Cryst.* A**30**, 180–184.

Sayre, D. (1980). *Phase extension and refinement using convolutional and related equation systems*. In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 271–286. New York, London: Plenum.

Schroeder, M. R. (1986). *Number Theory in Science and Communication*, 2nd ed. Berlin: Springer-Verlag.

Schwartz, L. (1965). *Mathematics for the Physical Sciences*. Paris: Hermann, and Reading: Addison-Wesley.

Schwartz, L. (1966). *Théorie des Distributions*. Paris: Hermann.

Schwarzenberger, R. L. E. (1980). *N-Dimensional Crystallography. Research Notes in Mathematics*, Vol. 41. London: Pitman.

Scott, W. R. (1964). *Group Theory*. Englewood Cliffs: Prentice-Hall. [Reprinted by Dover, New York, 1987.]

Shaffer, P. A. Jr, Schomaker, V. & Pauling, L. (1946*a*). *The use of punched cards in molecular structure determinations. I. Crystal structure calculations. J. Chem. Phys.* **14**, 648–658.

Shaffer, P. A. Jr, Schomaker, V. & Pauling, L. (1946*b*). *The use of punched cards in molecular structure determinations. II. Electron diffraction calculations. J. Chem. Phys.* **14**, 659–664.

Shannon, C. E. (1949). *Communication in the presence of noise. Proc. Inst. Radio Eng. NY*, **37**, 10–21.

Shenefelt, M. (1988). *Group Invariant Finite Fourier Transforms*. PhD thesis, Graduate Centre of the City University of New York.

Shmueli, U. & Weiss, G. H. (1985). *Exact joint probability distribution for centrosymmetric structure factors. Derivation and application to the* Σ_1 *relationship in the space group* $P\bar{1}$. *Acta Cryst.* A**41**, 401–408.

Shmueli, U. & Weiss, G. H. (1986). *Exact joint distribution of* $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ *and* $E_{\mathbf{h}+\mathbf{k}}$, *and the probability for the positive sign of the triple product in the space group* $P\bar{1}$. *Acta Cryst.* A**42**, 240–246.

Shmueli, U. & Weiss, G. H. (1987). *Exact random-walk models in crystallographic statistics. III. Distributions of* $|E|$ *for space groups of low symmetry. Acta Cryst.* A**43**, 93–98.

Shmueli, U. & Weiss, G. H. (1988). *Exact random-walk models in crystallographic statistics. IV. P.d.f.'s of* $|E|$ *allowing for atoms in special positions. Acta Cryst.* A**44**, 413–417.

Shmueli, U., Weiss, G. H. & Kiefer, J. E. (1985). *Exact random-walk models in crystallographic statistics. II. The bicentric distribution for the space group* $P\bar{1}$. *Acta Cryst.* A**41**, 55–59.

Shmueli, U., Weiss, G. H., Kiefer, J. E. & Wilson, A. J. C. (1984). *Exact random-walk models in crystallographic statistics. I. Space groups* $P\bar{1}$ *and* $P1$. *Acta Cryst.* A**40**, 651–660.

Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *The crystal structure of* L*S-threonine. J. Am. Chem. Soc.* **72**, 2328–2349.

Shoemaker, D. P. & Sly, W. G. (1961). *Computer programming strategy for crystallographic Fourier synthesis: program MIFR1. Acta Cryst.* **14**, 552.

Shohat, J. A. & Tamarkin, J. D. (1943). *The Problem of Moments. Mathematical Surveys*, No. 1. New York: American Mathematical Society.

Silverman, H. F. (1977). *An introduction to programming the Winograd Fourier transform algorithm (WFTA). IEEE Trans. Acoust. Speech Signal Process.* **25**, 152–165.

Singleton, R. C. (1969). *An algorithm for computing the mixed radix fast Fourier transform. IEEE Trans. AU*, **17**, 93–103.

Sneddon, I. N. (1951). *Fourier Transforms*. New York: McGraw-Hill.

Sneddon, I. N. (1972). *The Use of Integral Transforms*. New York: McGraw-Hill.

Sparks, R. A., Prosen, R. J., Kruse, F. H. & Trueblood, K. N. (1956). *Crystallographic calculations on the high-speed digital computer SWAC. Acta Cryst.* **9**, 350–358.

Sprecher, D. A. (1970). *Elements of Real Analysis*. New York: Academic Press. [Reprinted by Dover Publications, New York, 1987.]

Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination. A Practical Guide*. New York: Macmillan.

Suryan, G. (1957). *An analogue computer for double Fourier series summation for X-ray crystal-structure analysis. Acta Cryst.* **10**, 82–84.

Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. (1977). *A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. Acta Cryst.* A**33**, 800–804.

Swartzrauber, P. N. (1984). *FFT algorithms for vector computers. Parallel Comput.* **1**, 45–63.

Szegö, G. (1915). *Ein Grenzwertsatz uber die Toeplitzschen Determinanten einer reellen positiven Funktion. Math. Ann.* **76**, 490–503.

Szegö, G. (1920). *Beitrage zur Theorie der Toeplitzchen Formen (Erste Mitteilung). Math. Z.* **6**, 167–202.

Szegö, G. (1952). *On certain Hermitian forms associated with the Fourier series of a positive function. Comm. Sém. Mat., Univ. Lund* (Suppl. dedicated to Marcel Riesz), pp. 228–238.

Takano, T. (1977*a*). *Structure of myoglobin refined at 2.0 Å resolution. I. Crystallographic refinement of metmyoglobin from sperm whale. J. Mol. Biol.* **110**, 537–568.

Takano, T. (1977*b*). *Structure of myoglobin refined at 2.0 Å resolution. II. Structure of deoxymyoglobin from sperm whale. J. Mol. Biol.* **110**, 569–584.

Temperton, C. (1983*a*). *Self-sorting mixed-radix fast Fourier transforms. J. Comput. Phys.* **52**, 1–23.

Temperton, C. (1983*b*). *A note on the prime factor FFT algorithm. J. Comput. Phys.* **52**, 198–204.

Temperton, C. (1983*c*). *Fast mixed-radix real Fourier transforms. J. Comput. Phys.* **52**, 340–350.

Temperton, C. (1985). *Implementation of a self-sorting in place prime factor FFT algorithm. J. Comput. Phys.* **58**, 283–299.

Ten Eyck, L. F. (1973). *Crystallographic fast Fourier transforms. Acta Cryst*. A**29**, 183–191.

Ten Eyck, L. F. (1977). *Efficient structure factor calculation for large molecules by fast Fourier transform. Acta Cryst*. A**33**, 486–492.

Ten Eyck, L. F. (1985). *Fast Fourier transform calculation of electron density maps*. In *Diffraction Methods for Biological Macromolecules* (*Methods in Enzymology*, Vol. 115), edited by H. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 324–337. New York: Academic Press.

Titchmarsh, E. C. (1922). *Hankel transforms. Proc. Camb. Philos. Soc.* **21**, 463–473.

Titchmarsh, E. C. (1948). *Introduction to the Theory of Fourier Integrals*. Oxford: Clarendon Press.

Toeplitz, O. (1907). *Zur Theorie der quadratischen Formen von unendlichvielen Variablen. Nachr. der Kgl. Ges. Wiss. Göttingen, Math. Phys. Kl.* pp. 489–506.

Toeplitz, O. (1910). *Zur Transformation der Scharen bilinearer Formen von unendlichvielen Veränderlichen. Nachr. der Kgl. Ges. Wiss. Göttingen, Math. Phys. Kl.* pp. 110–115.

Toeplitz, O. (1911*a*). *Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. I. Teil: Theorie der L-formen. Math. Ann.* **70**, 351–376.

Toeplitz, O. (1911*b*). *Über die Fouriersche Entwicklung positiver Funktionen. Rend. Circ. Mat. Palermo*, **32**, 191–192.

Tolimieri, R. (1985). *The algebra of the finite Fourier transform and coding theory. Trans. Am. Math. Soc.* **287**, 253–273.

Tolimieri, R., An, M. & Lu, C. (1989). *Algorithms for Discrete Fourier Transform and Convolution*. New York: Springer-Verlag.

Tolstov, G. P. (1962). *Fourier Series*. Englewood Cliffs: Prentice-Hall.

Trèves, F. (1967). *Topological Vector Spaces, Distributions, and Kernels*. New York, London: Academic Press.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Cryst*. A**43**, 489–501.

Trueblood, K. N. (1956). *Symmetry transformations of general anisotropic temperature factors. Acta Cryst*. **9**, 359–361.

Truter, M. R. (1954). *Refinement of a non-centrosymmetrical structure: sodium nitrite. Acta Cryst*. **7**, 73–77.

Uhrich, M. L. (1969). *Fast Fourier transforms without sorting. IEEE Trans. AU*, **17**, 170–172.

Van der Pol, B. & Bremmer, H. (1955). *Operational Calculus*, 2nd ed. Cambridge University Press.

Vand, V. (1948). *Method of steepest descents: improved formula for X-ray analysis. Nature (London)*, **161**, 600–601.

Vand, V. (1951). *A simplified method of steepest descents. Acta Cryst*. **4**, 285–286.

Vand, V., Eiland, P. F. & Pepinsky, R. (1957). *Analytical representation of atomic scattering factors. Acta Cryst*. **10**, 303–306.

Wang, B. C. (1985). *Resolution of phase ambiguity in macromolecular crystallography*. In *Diffraction Methods for Biological Macromolecules* (*Methods in Enzymology*, Vol. 115), edited by H. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 90–112. New York: Academic Press.

Warren, B. & Bragg, W. L. (1929). *The structure of diopside, $CaMg(SiO_3)_2$. Z. Kristallogr.* **69**, 168–193.

Warren, B. E. (1969). *X-ray Diffraction*. Reading: Addison-Wesley.

Warren, B. E. & Gingrich, N. S. (1934). *Fourier integral analysis of X-ray powder patterns. Phys. Rev*. **46**, 368–372.

Waser, J. (1955*a*). *Symmetry relations between structure factors. Acta Cryst*. **8**, 595.

Waser, J. (1955*b*). *The anisotropic temperature factor in triclinic coordinates. Acta Cryst*. **8**, 731.

Waser, J. & Schomaker, V. (1953). *The Fourier inversion of diffraction data. Rev. Mod. Phys.* **25**, 671–690.

Watson, G. L. (1970). *Integral Quadratic Forms*. Cambridge University Press.

Watson, G. N. (1944). *A Treatise on the Theory of Bessel Functions*, 2nd ed. Cambridge University Press.

Wells, M. (1965). *Computational aspects of space-group symmetry. Acta Cryst*. **19**, 173–179.

Weyl, H. (1931). *The Theory of Groups and Quantum Mechanics*. New York: Dutton. [Reprinted by Dover Publications, New York, 1950.]

Whittaker, E. J. W. (1948). *The evaluation of Fourier transforms by a Fourier synthesis method. Acta Cryst*. **1**, 165–167.

Whittaker, E. T. (1915). *On the functions which are represented by the expansions of the interpolation-theory. Proc. R. Soc. (Edinburgh)*, **35**, 181–194.

Whittaker, E. T. (1928). *Oliver Heaviside. Bull. Calcutta Math. Soc.* **20**, 199–220. [Reprinted in Moore (1971).]

Whittaker, E. T. & Robinson, G. (1944). *The Calculus of Observations*. London: Blackie.

Whittaker, E. T. & Watson, G. N. (1927). *A Course of Modern Analysis*, 4th ed. Cambridge University Press.

Widom, H. (1960). *A theorem on translation kernels in n dimensions. Trans. Am. Math. Soc.* **94**, 170–180.

Widom, H. (1965). *Toeplitz matrices*. In *Studies in Real and Complex Analysis*, edited by I. I. Hirschmann Jr, pp. 179–209. *MAA Studies in Mathematics*, Vol. 3. Englewood Cliffs: Prentice-Hall.

Widom, H. (1975). *Asymptotic inversion of convolution operators. Publ. Math. IHES*, **44**, 191–240.

Wiener, N. (1933). *The Fourier Integral and Certain of its Applications*. Cambridge University Press. [Reprinted by Dover Publications, New York, 1959.]

Wilkins, S. W., Varghese, J. N. & Lehmann, M. S. (1983). *Statistical geometry. I. A self-consistent approach to the crystallographic inversion problem based on information theory. Acta Cryst.* A**39**, 47–60.

Winograd, S. (1976). *On computing the discrete Fourier transform. Proc. Natl Acad. Sci. USA*, **73**, 1005–1006.

Winograd, S. (1977). *Some bilinear forms whose multiplicative complexity depends on the field of constants. Math. Syst. Theor.* **10**, 169–180.

Winograd, S. (1978). *On computing the discrete Fourier transform. Math. Comput.* **32**, 175–199.

Winograd, S. (1980). *Arithmetic Complexity of Computations*. CBMS-NST Regional Conf. Series Appl. Math, Publ. No. 33. Philadelphia: SIAM Publications.

Wolf, J. A. (1967). *Spaces of Constant Curvature.* New York: McGraw-Hill.

Wondratschek, H. (2005). *Introduction to space-group symmetry.* In *International Tables for Crystallography*, Vol. A. *Space-Group Symmetry*, edited by Th. Hahn, Part 8. Heidelberg: Springer.

Yosida, K. (1965). *Functional Analysis.* Berlin: Springer-Verlag.

Zachariasen, W. H. (1945). *Theory of X-ray Diffraction in Crystals.* New York: John Wiley.

Zassenhaus, H. (1948). *Über eine Algorithmus zur Bestimmung der Raumgruppen. Commun. Helv. Math.* **21**, 117–141.

Zemanian, A. H. (1965). *Distribution Theory and Transform Analysis.* New York: McGraw-Hill.

Zemanian, A. H. (1968). *Generalised Integral Transformations.* New York: Interscience.

Zygmund, A. (1959). *Trigonometric Series*, Vols. 1 and 2. Cambridge University Press.

Zygmund, A. (1976). *Notes on the history of Fourier series*. In *Studies in Harmonic Analysis,* edited by J. M. Ash, pp. 1–19. *MAA Studies in Mathematics*, Vol. 13. The Mathematical Association of America.

International Tables for Crystallography (2010). Vol. B, Chapter 1.4, pp. 114–174.

1.4. Symmetry in reciprocal space

By U. Shmueli

With Appendix A1.4.2 by U. Shmueli, S. R. Hall and R. W. Grosse-Kunstleve

1.4.1. Introduction

Crystallographic symmetry, as reflected in functions on reciprocal space, can be considered from two complementary points of view.

(1) One can assume the existence of a certain permissible symmetry of the density function of crystalline (scattering) matter, a function which due to its three-dimensional periodicity can be expanded in a triple Fourier series (*e.g.* Bragg, 1966), and inquire about the effects of this symmetry on the Fourier coefficients – the structure factors. Since there exists a one-to-one correspondence between the triplets of summation indices in the Fourier expansion and vectors in the reciprocal lattice (Ewald, 1921), the above approach leads to consequences of the symmetry of the density function which are relevant to the representation of its Fourier image in reciprocal space. The symmetry properties of these Fourier coefficients, which are closely related to the crystallographic experiment, can then be readily established.

This traditional approach, the essentials of which are the basis of Sections 4.5–4.7 of Volume I (*IT* I, 1952), and which was further developed in the works of Buerger (1949, 1960), Waser (1955), Bertaut (1964) and Wells (1965), is one of the cornerstones of crystallographic practice and will be followed in the present chapter, as far as the basic principles are concerned.

(2) The alternative approach, proposed by Bienenstock & Ewald (1962), also presumes a periodic density function in crystal space and its Fourier expansion associated with the reciprocal. However, the argument starts from the Fourier coefficients, taken as a discrete set of complex functions, and linear transformations are sought which leave the magnitudes of these functions unchanged; the variables on which these transformations operate are h, k, l and φ – the Fourier summation indices (*i.e.*, components of a reciprocal-lattice vector) and the phase of the Fourier coefficient, respectively. These transformations, or the groups they constitute, are then interpreted in terms of the symmetry of the density function in direct space. This direct analysis of symmetry in reciprocal space will also be discussed.

We start the next section with a brief discussion of the point-group symmetries of associated direct and reciprocal lattices. The weighted reciprocal lattice is then briefly introduced and the relation between the values of the weight function at symmetry-related points of the weighted reciprocal lattice is discussed in terms of the Fourier expansion of a periodic function in crystal space. The remaining part of Section 1.4.2 is devoted to the formulation of the Fourier series and its coefficients (values of the weight function) in terms of space-group-specific symmetry factors. Section 1.4.3 then explains the basis for an automated generation of simplified geometrical structure-factor formulae, which are presented for all the two- and three-dimensional space groups in Appendix 1.4.3. This is a revised version of the structure-factor tables given in Sections 4.5–4.7 of Volume I (*IT* I, 1952). Appendix 1.4.4 contains a reciprocal-space representation of the 230 crystallographic space groups and some explanatory material related to these space-group tables is given in Section 1.4.4; the latter are interpreted in terms of the two viewpoints discussed above. The tabular material given in this chapter is compatible with the direct-space symmetry tables given in Volume A (*IT* A, 1983) with regard to the space-group settings and choices of the origin.

Most of the tabular material, the new symmetry-factor tables in Appendix 1.4.3 and the space-group tables in Appendix 1.4.4 have been generated by computer with the aid of a combination of numeric and symbolic programming techniques. The algorithm underlying this procedure is briefly summarized in Appendix 1.4.1. Appendix 1.4.2 deals with computer-adapted space-group symbols, including the set of symbols that were used in the preparation of the present tables.

Computer generation of symmetry information is not new. However, we can quote the Bilbao Crystallographic Server (*e.g.* Aroyo *et al.*, 2006) as a rich source of symmetry information which is readily accessible from the Internet.

1.4.2. Effects of symmetry on the Fourier image of the crystal

1.4.2.1. Point-group symmetry of the reciprocal lattice

Regarding the reciprocal lattice as a collection of points generated from a given direct lattice, it is fairly easy to see that each of the two associated lattices must have the same point-group symmetry. The set of all the rotations that bring the direct lattice into self-coincidence can be thought of as interchanging equivalent families of lattice planes in all the permissible manners. A family of lattice planes in the direct lattice is characterized by a common normal and a certain interplanar distance, and these two characteristics uniquely define the direction and magnitude, respectively, of a vector in the reciprocal lattice, as well as the lattice line associated with this vector and passing through the origin. It follows that any symmetry operation on the direct lattice must also bring the reciprocal lattice into self-coincidence, *i.e.* it must also be a symmetry operation on the reciprocal lattice. The roles of direct and reciprocal lattices in the above argument can of course be interchanged without affecting the conclusion.

The above elementary considerations recall that for any point group (not necessarily the full point group of a lattice), the operations which leave the lattice unchanged must also leave unchanged its associated reciprocal. This equivalence of point-group symmetries of the associated direct and reciprocal lattices is fundamental to crystallographic symmetry in reciprocal space, in both points of view mentioned in Section 1.4.1.

With regard to the effect of any given point-group operation on each of the two associated lattices, we recall that:

(i) If $\boldsymbol{P}$ is a point-group rotation operator acting on the direct lattice (*e.g.* by rotation through the angle α about a given axis), the effect of this rotation on the associated reciprocal lattice is that of applying the inverse rotation operator, $\boldsymbol{P}^{-1}$ (*i.e.* rotation through $-\alpha$ about a direction parallel to the direct axis); this is readily found from the requirement that the scalar product $\mathbf{h}^T\mathbf{r}_\mathrm{L}$, where $\mathbf{h}$ and $\mathbf{r}_\mathrm{L}$ are vectors in the reciprocal and direct lattices, respectively, remains invariant under the application of a point-group operation to the crystal.

(ii) If our matrix representation of the rotation operator is such that the point-group operation is applied to the direct-lattice (column) vector by *premultiplying* it with the matrix $\boldsymbol{P}$, the corresponding operation on the reciprocal lattice is applied by *postmultiplying* the (row) vector $\mathbf{h}^T$ with the point-group rotation matrix. We can thus write, *e.g.*, $\mathbf{h}^T\mathbf{r}_\mathrm{L} = (\mathbf{h}^T\boldsymbol{P}^{-1})(\boldsymbol{P}\mathbf{r}_\mathrm{L}) = [(\boldsymbol{P}^{-1})^T\mathbf{h}]^T(\boldsymbol{P}\mathbf{r}_\mathrm{L})$. Note, however, that the orthogonality relation-

Copyright © 2010 International Union of Crystallography

ship: $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$ is not satisfied if $\boldsymbol{P}$ is referred to some oblique crystal systems, higher than the orthorhombic.

Detailed descriptions of the 32 crystallographic point groups are presented in the crystallographic and other literature; their complete tabulation is given in Part 10 of Volume A (*IT* A, 2005).

1.4.2.2. Relationship between structure factors at symmetry-related points of the reciprocal lattice

Of main interest in the context of the present chapter are symmetry relationships that concern the values of a function defined at the points of the reciprocal lattice. Such functions, of crystallographic interest, are Fourier-transform representations of direct-space functions that have the periodicity of the crystal, the structure factor as a Fourier transform of the electron-density function being a representative example (see *e.g.* Lipson & Taylor, 1958). The value of such a function, attached to a reciprocal-lattice point, is called the weight of this point and the set of all such weighted points is often termed the weighted reciprocal lattice. This section deals with a fundamental relationship between functions (weights) associated with reciprocal-lattice points, which are related by point-group symmetry, the weights here considered being the structure factors of Bragg reflections (*cf.* Chapter 1.2).

The electron density, an example of a three-dimensional periodic function with the periodicity of the crystal, can be represented by the Fourier series

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}^T \mathbf{r}), \tag{1.4.2.1}$$

where $\mathbf{h}$ is a reciprocal-lattice vector, V is the volume of the (direct) unit cell, $F(\mathbf{h})$ is the structure factor at the point $\mathbf{h}$ and $\mathbf{r}$ is a position vector of a point in direct space, at which the density is given. The summation in (1.4.2.1) extends over all the reciprocal lattice.

Let $\mathbf{r}' = \boldsymbol{P}\mathbf{r} + \mathbf{t}$ be a space-group operation on the crystal, where $\boldsymbol{P}$ and $\mathbf{t}$ are its rotation and translation parts, respectively, and $\boldsymbol{P}$ must therefore be a point-group operator. We then have, by definition, $\rho(\mathbf{r}) = \rho(\boldsymbol{P}\mathbf{r} + \mathbf{t})$ and the Fourier representation of the electron density, at the equivalent position $\boldsymbol{P}\mathbf{r} + \mathbf{t}$, is given by

$$\begin{aligned}\rho(\boldsymbol{P}\mathbf{r} + \mathbf{t}) &= \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h}) \exp[-2\pi i \mathbf{h}^T(\boldsymbol{P}\mathbf{r} + \mathbf{t})] \\ &= \frac{1}{V}\sum_{\mathbf{h}} [F(\mathbf{h}) \exp(-2\pi i \mathbf{h}^T \mathbf{t})] \\ &\quad \times \exp[-2\pi i (\boldsymbol{P}^T\mathbf{h})^T \mathbf{r}],\end{aligned} \tag{1.4.2.2}$$

noting that $\mathbf{h}^T\boldsymbol{P} = (\boldsymbol{P}^T\mathbf{h})^T$. Since $\boldsymbol{P}$ is a point-group operator, the vectors $\boldsymbol{P}^T\mathbf{h}$ in (1.4.2.2) must range over all the reciprocal lattice and a comparison of the functional forms of the equivalent expansions (1.4.2.1) and (1.4.2.2) shows that the coefficients of the exponentials $\exp[-2\pi i(\boldsymbol{P}^T\mathbf{h})^T\mathbf{r}]$ in (1.4.2.2) must be the structure factors at the points $\boldsymbol{P}^T\mathbf{h}$ in the reciprocal lattice. Thus

$$F(\boldsymbol{P}^T\mathbf{h}) = F(\mathbf{h}) \exp(-2\pi i \mathbf{h}^T \mathbf{t}), \tag{1.4.2.3}$$

wherefrom it follows that the magnitudes of the structure factors at $\mathbf{h}$ and $\boldsymbol{P}^T\mathbf{h}$ are the same:

$$|F(\boldsymbol{P}^T\mathbf{h})| = |F(\mathbf{h})|, \tag{1.4.2.4}$$

and their phases are related by

$$\varphi(\boldsymbol{P}^T\mathbf{h}) = \varphi(\mathbf{h}) - 2\pi \mathbf{h}^T\mathbf{t}. \tag{1.4.2.5}$$

The relationship (1.4.2.3) between structure factors of symmetry-related reflections was first derived by Waser (1955), starting from a representation of the structure factor as a Fourier transform of the electron-density function.

It follows that an application of a point-group transformation to the (weighted) reciprocal lattice leaves the moduli of the structure factors unchanged. The distribution of diffracted intensities obeys, in fact, the same point-group symmetry as that of the crystal. If, however, anomalous dispersion is negligibly small, and the point group of the crystal is noncentrosymmetric, the apparent symmetry of the diffraction pattern will also contain a false centre of symmetry and, of course, all the additional elements generated by the inclusion of this centre. Under these circumstances, the diffraction pattern from a single crystal may belong to one of the eleven centrosymmetric point groups, known as Laue groups (*IT* I, 1952).

According to equation (1.4.2.5), the phases of the structure factors of symmetry-related reflections differ, in the general case, by a phase shift that depends on the translation part of the space-group operation involved. Only when the space group is symmorphic, *i.e.* it contains no translations other than those of the Bravais lattice, will the distribution of the phases obey the point-group symmetry of the crystal. These phase shifts are considered in detail in Section 1.4.4 where their tabulation is presented and the alternative interpretation (Bienenstock & Ewald, 1962) of symmetry in reciprocal space, mentioned in Section 1.4.1, is given.

Equation (1.4.2.3) can be usefully applied to a classification of all the general systematic absences or – as defined in the space-group tables in the main editions of *IT* (1935, 1952, 1983, 1987, 1992, 1995, 2002) – general conditions for possible reflections. These systematic absences are associated with special positions in the reciprocal lattice – special with respect to the point-group operations $\boldsymbol{P}$ appearing in the relevant relationships. If, in a given relationship, we have $\boldsymbol{P}^T\mathbf{h} = \mathbf{h}$, equation (1.4.2.3) reduces to

$$F(\mathbf{h}) = F(\mathbf{h}) \exp(-2\pi i \mathbf{h}^T \mathbf{t}). \tag{1.4.2.6}$$

Of course, $F(\mathbf{h})$ may then be nonzero only if $\cos(2\pi \mathbf{h}^T\mathbf{t})$ equals unity, or the scalar product $\mathbf{h}^T\mathbf{t}$ is an integer. This well known result leads to a ready determination of lattice absences, as well as those produced by screw-axis and glide-plane translations, and is routinely employed in crystallographic computing. An exhaustive classification of the general conditions for possible reflections is given in the space-group tables (*IT*, 1952, 1983). It should be noted that since the axes of rotation and planes of reflection in the reciprocal lattice are parallel to the corresponding elements in the direct lattice (Buerger, 1960), the component of $\mathbf{t}$ that depends on the location of the corresponding space-group symmetry element in direct space does not contribute to the scalar product $\mathbf{h}^T\mathbf{t}$ in (1.4.2.6), and it is only the intrinsic part of the translation $\mathbf{t}$ (*IT* A, 1983) that usually matters.

It may, however, be of interest to note that some screw axes in direct space cannot give rise to any systematic absences. For example, the general Wyckoff position No. (10) in the space group $Pa\bar{3}$ (No. 205) (*IT* A, 1983) has the coordinates $\bar{y}, \frac{1}{2}+z, \frac{1}{2}-x$, and corresponds to the space-group operation

$$(\boldsymbol{P}, \mathbf{t}) = (\boldsymbol{P}, \mathbf{t}_i + \mathbf{t}_l) = \left[\begin{pmatrix} 0 & \bar{1} & 0 \\ 0 & 0 & 1 \\ \bar{1} & 0 & 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{6} \\ \frac{1}{6} \end{pmatrix}\right], \tag{1.4.2.7}$$

where $\mathbf{t}_i$ and $\mathbf{t}_l$ are the intrinsic and location-dependent components of the translation part $\mathbf{t}$, and are parallel and perpendicular, respectively, to the threefold axis of rotation represented by the matrix $\boldsymbol{P}$ in (1.4.2.7) (*IT* A, 1983; Shmueli, 1984). This is clearly a threefold screw axis, parallel to $[\bar{1}\bar{1}1]$. The reciprocal-lattice vectors which remain unchanged, when postmultiplied by $\boldsymbol{P}$ (or premultiplied by its transpose), have the form: $\mathbf{h}^T = (h\bar{h}\bar{h})$; this is the special position for the present example. We see that (i) $\mathbf{h}^T\mathbf{t}_l = 0$, as expected, and (ii) $\mathbf{h}^T\mathbf{t}_i = -h$. Since the scalar product $\mathbf{h}^T\mathbf{t}$ is an integer, there are no values of index h for which the structure factor $F(h\bar{h}\bar{h})$ must be absent.

Other approaches to systematically absent reflections include a direct inspection of the structure-factor equation (Lipson & Cochran, 1966), which is of considerable didactical value, and the utilization of transformation properties of direct and reciprocal base vectors and lattice-point coordinates (Buerger, 1942).

Finally, the relationship between the phases of symmetry-related reflections, given by (1.4.2.5), is of fundamental as well as practical importance in the theories and techniques of crystal structure determination which operate in reciprocal space (Part 2 of this volume).

1.4.2.3. Symmetry factors for space-group-specific Fourier summations

The weighted reciprocal lattice, with weights taken as the structure factors, is synonymous with the discrete space of the coefficients of a Fourier expansion of the electron density, or the Fourier space (*F* space) of the latter. Accordingly, the asymmetric unit of the Fourier space can be defined as the subset of structure factors within which the relationship (1.4.2.3) does not hold – except at special positions in the reciprocal lattice. If the point group of the crystal is of order g, this is also the order of the corresponding factor-group representation of the space group (*IT* A, 1983) and there exist g relationships of the form of (1.4.2.3):

$$F(\boldsymbol{P}_s^T\mathbf{h}) = F(\mathbf{h})\exp(-2\pi i\mathbf{h}^T\mathbf{t}_s). \tag{1.4.2.8}$$

We can thus decompose the summation in (1.4.2.1) into g sums, each extending over an asymmetric unit of the *F* space. It must be kept in mind, however, that some classes of reciprocal-lattice vectors may be common to more than one asymmetric unit, and thus each reciprocal-lattice point will be assigned an occupancy factor, denoted by $q(\mathbf{h})$, such that $q(\mathbf{h}) = 1$ for a general position and $q(\mathbf{h}) = 1/m(\mathbf{h})$ for a special one, where $m(\mathbf{h})$ is the multiplicity – or the order of the point group that leaves $\mathbf{h}$ unchanged. Equation (1.4.2.1) can now be rewritten as

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{s=1}^{g}\sum_{\mathbf{h}_a} q(\mathbf{h}_a)F(\boldsymbol{P}_s^T\mathbf{h}_a)\exp[-2\pi i(\boldsymbol{P}_s^T\mathbf{h}_a)^T\mathbf{r}], \tag{1.4.2.9}$$

where the inner summation in (1.4.2.9) extends over the reference asymmetric unit of the Fourier space, which is associated with the identity operation of the space group. Substituting from (1.4.2.8) for $F(\boldsymbol{P}_s^T\mathbf{h}_a)$, and interchanging the order of the summations in (1.4.2.9), we obtain

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{h}_a} q(\mathbf{h}_a)F(\mathbf{h}_a)\sum_{s=1}^{g}\exp[-2\pi i\mathbf{h}_s^T(\boldsymbol{P}_s\mathbf{r}+\mathbf{t}_s)] \tag{1.4.2.10}$$

$$= \frac{1}{V}\sum_{\mathbf{h}_a} q(\mathbf{h}_a)F(\mathbf{h}_a)[A(\mathbf{h}_a) - iB(\mathbf{h}_a)], \tag{1.4.2.11}$$

where

$$A(\mathbf{h}) = \sum_{s=1}^{g}\cos[2\pi\mathbf{h}^T(\boldsymbol{P}_s\mathbf{r}+\mathbf{t}_s)] \tag{1.4.2.12}$$

and

$$B(\mathbf{h}) = \sum_{s=1}^{g}\sin[2\pi\mathbf{h}^T(\boldsymbol{P}_s\mathbf{r}+\mathbf{t}_s)]. \tag{1.4.2.13}$$

The symmetry factors A and B are well known as geometric or trigonometric structure factors and a considerable part of Volume I of *IT* (1952) is dedicated to their tabulation. Their formal association with the structure factor – following from direct-space arguments – is closely related to that shown in equation (1.4.2.11) (see Section 1.4.2.4). Simplified trigonometric expressions for A and B are given in Tables A1.4.3.1–A1.4.3.7 in Appendix 1.4.3 for all the two- and three-dimensional crystallographic space groups, and for all the parities of hkl for which A and B assume different functional forms. These expressions are there given for general reflections and can also be used for special ones, provided the occupancy factors $q(\mathbf{h})$ have been properly accounted for.

Equation (1.4.2.11) is quite general and can, of course, be applied to noncentrosymmetric Fourier summations, without neglect of dispersion. Further simplifications are obtained in the centrosymmetric case, when the space-group origin is chosen at a centre of symmetry, and in the noncentrosymmetric case, when dispersion is neglected. In each of the latter two cases the summation over $\mathbf{h}_a$ is restricted to reciprocal-lattice vectors that are not related by real or apparent inversion (denoted by $\mathbf{h}_a > 0$), and we obtain

$$\rho(\mathbf{r}) = \frac{2}{V}\sum_{\mathbf{h}_a>0} q(\mathbf{h}_a)F(\mathbf{h}_a)A(\mathbf{h}_a) \tag{1.4.2.14}$$

and

$$\rho(\mathbf{r}) = \frac{2}{V}\sum_{\mathbf{h}_a>0} q(\mathbf{h}_a)|F(\mathbf{h}_a)|[A(\mathbf{h}_a)\cos\varphi(\mathbf{h}_a) + B(\mathbf{h}_a)\sin\varphi(\mathbf{h}_a)] \tag{1.4.2.15}$$

for the dispersionless centrosymmetric and noncentrosymmetric cases, respectively.

1.4.2.4. Symmetry factors for space-group-specific structure-factor formulae

The explicit dependence of structure-factor summations on the space-group symmetry of the crystal can also be expressed in terms of symmetry factors, in an analogous manner to that described for the electron density in the previous section. It must be pointed out that while the above treatment only presumes that the electron density can be represented by a three-dimensional Fourier series, the present one is restricted by the assumption that the atoms are isotropic with regard to their motion and shape (*cf.* Chapter 1.2).

Under the above assumptions, *i.e.* for isotropically vibrating spherical atoms, the structure factor can be written as

$$F(\mathbf{h}) = \sum_{j=1}^{N} f_j\exp(2\pi i\mathbf{h}^T\mathbf{r}_j), \tag{1.4.2.16}$$

where $\mathbf{h}^T = (hkl)$ is the diffraction vector, N is the number of atoms in the unit cell, f_j is the atomic scattering factor including

its temperature factor and depending on the magnitude of **h** only, and $\mathbf{r}_j$ is the position vector of the jth atom referred to the origin of the unit cell.

If the crystal belongs to a point group of order m_p and the multiplicity of its Bravais lattice is m_L, there are $g' = m_p \times m_L$ general equivalent positions in the unit cell of the space group (*IT* A, 1983). We can thus rewrite (1.4.2.16), grouping the contributions of the symmetry-related atoms, as

$$F(\mathbf{h}) = \sum_j f_j \sum_{s=1}^{g'} \exp[2\pi i \mathbf{h}^T (\boldsymbol{P}_s \mathbf{r} + \mathbf{t}_s)], \tag{1.4.2.17}$$

where $\boldsymbol{P}_s$ and $\mathbf{t}_s$ are the rotation and translation parts of the sth space-group operation respectively. The inner summation in (1.4.2.17) contains the dependence of the structure factor of reflection **h** on the space-group symmetry of the crystal and is known as the (complex) geometric or trigonometric structure factor.

Equation (1.4.2.17) can be rewritten as

$$F(\mathbf{h}) = \sum_j f_j [A_j(\mathbf{h}) + iB_j(\mathbf{h})], \tag{1.4.2.18}$$

where

$$A_j(\mathbf{h}) = \sum_{s=1}^{g'} \cos[2\pi \mathbf{h}^T (\boldsymbol{P}_s \mathbf{r}_j + \mathbf{t}_s)] \tag{1.4.2.19}$$

and

$$B_j(\mathbf{h}) = \sum_{s=1}^{g'} \sin[2\pi \mathbf{h}^T (\boldsymbol{P}_s \mathbf{r}_j + \mathbf{t}_s)] \tag{1.4.2.20}$$

are the real and imaginary parts of the trigonometric structure factor. Equations (1.4.2.19) and (1.4.2.20) are *mathematically* identical to equations (1.4.2.11) and (1.4.2.12), respectively, apart from the numerical coefficients which appear in the expressions for A and B, for space groups with centred lattices: while only the order of the point group need be considered in connection with the Fourier expansion of the electron density (see above), the multiplicity of the Bravais lattice must of course appear in (1.4.2.19) and (1.4.2.20).

Analogous functional forms are arrived at by considerations of symmetry in direct and reciprocal spaces. These quantities are therefore convenient representations of crystallographic symmetry in its interaction with the diffraction experiment and have been indispensable in all of the early crystallographic computing related to structure determination. Their applications to modern crystallographic computing have been largely superseded by fast Fourier techniques, in reciprocal space, and by direct use of matrix and vector representations of space-group operators, in direct space, especially in cases of low space-group symmetry. It should be noted, however, that the degree of simplification of the trigonometric structure factors generally increases with increasing symmetry (see, *e.g.*, Section 1.4.3), and the gain of computing efficiency becomes significant when problems involving high symmetries are treated with this 'old-fashioned' tool. Analytic expressions for the trigonometric structure factors are of course indispensable in studies in which the knowledge of the functional form of the structure factor is required [*e.g.* in theories of structure-factor statistics and direct methods of phase determination (see Chapters 2.1 and 2.2)].

Equations (1.4.2.19) and (1.4.2.20) are simple but their expansion and simplification for all the space groups and relevant *hkl* subsets can be an extremely tedious undertaking when carried out in the conventional manner. As shown below, this process has been automated by a suitable combination of symbolic and numeric high-level programming procedures.

1.4.3. Structure-factor tables

1.4.3.1. Some general remarks

This section is a revised version of the structure-factor tables contained in Sections 4.5 through 4.7 of Volume I (*IT* I, 1952). As in the previous edition, it is intended to present a comprehensive list of explicit expressions for the real and the imaginary parts of the trigonometric structure factor, for all the 17 plane groups and the 230 space groups, and for the *hkl* subsets for which the trigonometric structure factor assumes different functional forms. The tables given here are also confined to the case of general Wyckoff positions (*IT* I, 1952). However, the expressions are presented in a much more concise symbolic form and are amenable to computation just like the explicit trigonometric expressions in Volume I (*IT* I, 1952). The present tabulation is based on equations (1.4.2.19) and (1.4.2.20), *i.e.* the numerical coefficients in A and B which appear in Tables A1.4.3.1–A1.4.3.7 in Appendix 1.4.3 are appropriate to space-group-specific structure-factor formulae. The functional form of A and B is, however, the same when applied to Fourier summations (see Section 1.4.2.3).

1.4.3.2. Preparation of the structure-factor tables

The lists of the coordinates of the general equivalent positions, presented in *IT* A (1983), as well as in earlier editions of the *Tables*, are sufficient for the expansion of the summations in (1.4.2.19) and (1.4.2.20) and the simplification of the resulting expressions can be performed using straightforward algebra and trigonometry (see, *e.g.*, *IT* I, 1952). As mentioned above, the preparation of the present structure-factor tables has been automated and its stages can be summarized as follows:

(i) Generation of the coordinates of the general positions, starting from a computer-adapted space-group symbol (Shmueli, 1984).

(ii) Formation of the scalar products, appearing in (1.4.2.19) and (1.4.2.20), and their separation into components depending on the rotation and translation parts of the space-group operations:

$$\mathbf{h}^T(\boldsymbol{P}_s, \mathbf{t}_s)\mathbf{r} = \mathbf{h}^T \boldsymbol{P}_s \mathbf{r} + \mathbf{h}^T \mathbf{t}_s \tag{1.4.3.1}$$

for the space groups which are not associated with a unique axis; the left-hand side of (1.4.3.1) is separated into contributions of the relevant plane group and unique axis for the remaining space groups.

(iii) Analysis of the translation-dependent parts of the scalar products and automatic determination of all the parities of *hkl* for which A and B must be computed and simplified.

(iv) Expansion of equations (1.4.2.19) and (1.4.2.20) and their reduction to trigonometric expressions comparable to those given in the structure-factor tables in Volume I of *IT* (1952).

(v) Representation of the results in terms of a small number of building blocks, of which the expressions were found to be composed. These representations are described in Section 1.4.3.3.

All the stages outlined above were carried out with suitably designed computer programs, written in numerically and symbolically oriented languages. A brief summary of the underlying algorithms is presented in Appendix 1.4.1. The computer-adapted space-group symbols used in these computations are described in Section A1.4.2.2 and presented in Table A1.4.2.1.

1.4.3.3. Symbolic representation of A and B

We shall first discuss the symbols for the space groups that are not associated with a unique axis. These comprise the triclinic, orthorhombic and cubic space groups. The symbols are also used for the seven rhombohedral space groups which are referred to rhombohedral axes (*IT* I, 1952; *IT* A, 1983).

The abbreviation of triple products of trigonometric functions such as, *e.g.*, denoting $\cos(2\pi hx)\sin(2\pi ky)\cos(2\pi lz)$ by csc, is well known (*IT* I, 1952), and can be conveniently used in representing *A* and *B* for triclinic and orthorhombic space groups. However, the simplified expressions for *A* and *B* in space groups of higher symmetry also possess a high degree of regularity, as is apparent from an examination of the structure-factor tables in Volume I (*IT* I, 1952), and as confirmed by the preparation of the present tables. An example, illustrating this for the cubic system, is given below.

The trigonometric structure factor for the space group $Pm\bar{3}$ (No. 200) is given by

$$\begin{aligned} A = 8[&\cos(2\pi hx)\cos(2\pi ky)\cos(2\pi lz) \\ &+ \cos(2\pi hy)\cos(2\pi kz)\cos(2\pi lx) \\ &+ \cos(2\pi hz)\cos(2\pi kx)\cos(2\pi ly)], \end{aligned} \tag{1.4.3.2}$$

and the sum of the above nine-function block and the following one:

$$\begin{aligned} 8[&\cos(2\pi hx)\cos(2\pi kz)\cos(2\pi ly) \\ &+ \cos(2\pi hz)\cos(2\pi ky)\cos(2\pi lx) \\ &+ \cos(2\pi hy)\cos(2\pi kx)\cos(2\pi lz)] \end{aligned} \tag{1.4.3.3}$$

is the trigonometric structure factor for the space group $Pm\bar{3}m$ (No. 221, *IT* I, 1952, *IT* A, 1983). It is obvious that the only difference between the nine-function blocks in (1.4.3.2) and (1.4.3.3) is that the permutation of the coordinates *xyz* is *cyclic* or *even* in (1.4.3.2), while it is *noncyclic* or *odd* in (1.4.3.3).

It was observed during the generation of the present tables that the expressions for *A* and *B* for all the cubic space groups, and all the relevant *hkl* subsets, can be represented in terms of such 'even' and 'odd' nine-function blocks. Moreover, it was found that the order of the trigonometric functions in each such block remains the same in each of its three terms (triple products). This is not surprising since each of the above space groups contains threefold axes of rotation along [111] and related directions, and such permutations of *xyz* for fixed *hkl* (or *vice versa*) are expected. It was therefore possible to introduce two permutation operators and represent *A* and *B* in terms of the following two basic blocks:

$$\begin{aligned} \mathrm{Epqr} = {}&\mathrm{p}(2\pi hx)\mathrm{q}(2\pi ky)\mathrm{r}(2\pi lz) + \mathrm{p}(2\pi hy)\mathrm{q}(2\pi kz)\mathrm{r}(2\pi lx) \\ &+ \mathrm{p}(2\pi hz)\mathrm{q}(2\pi kx)\mathrm{r}(2\pi ly) \end{aligned} \tag{1.4.3.4}$$

and

$$\begin{aligned} \mathrm{Opqr} = {}&\mathrm{p}(2\pi hx)\mathrm{q}(2\pi kz)\mathrm{r}(2\pi ly) + \mathrm{p}(2\pi hz)\mathrm{q}(2\pi ky)\mathrm{r}(2\pi lx) \\ &+ \mathrm{p}(2\pi hy)\mathrm{q}(2\pi kx)\mathrm{r}(2\pi lz), \end{aligned} \tag{1.4.3.5}$$

where each of p, q and r can be a sine or a cosine, and appears at the same position in each of the three terms of a block. The capital prefixes E and O were chosen to represent even and odd permutations of the coordinates *xyz*, respectively.

For example, the trigonometric structure factor for the space group $Pa\bar{3}$ (No. 205, *IT* I, 1952, *IT* A, 1983) can now be tabulated as follows:

A	*B*	$h+k$	$k+l$	$h+l$
8Eccc	0	even	even	even
−8Ecss	0	even	odd	odd
−8Escs	0	odd	even	odd
−8Essc	0	odd	odd	even

(*cf.* Table A1.4.3.7), where the sines and cosines are abbreviated by s and c, respectively. It is interesting to note that the only maximal non-isomorphic subgroup of $Pa\bar{3}$, not containing a threefold axis, is the orthorhombic *Pbca* (see *IT* A, 1983, p. 621), and this group–subgroup relationship is reflected in the functional forms of the trigonometric structure factors; the representation of *A* and *B* for *Pbca* is in fact analogous to that of $Pa\bar{3}$, including the parities of *hkl* and the corresponding forms of the triple products, except that the prefix E – associated with the threefold rotation – is absent from *Pbca*. The expression for *A* for the space group $Pm\bar{3}m$ [the sum of (1.4.3.2) and (1.4.3.3)] now simply reads: $A = 8(\mathrm{Eccc} + \mathrm{Occc})$.

As pointed out above, the permutation operators also apply to rhombohedral space groups that are referred to rhombohedral axes (Table A1.4.3.6), and the corresponding expressions for *R*3 and $R\bar{3}$ bear the same relationship to those for *P*1 and $P\bar{1}$ (Table A1.4.3.2), respectively, as that shown above for the related $Pa\bar{3}$ and *Pbca*.

When in any given standard space-group setting one of the coordinate axes is parallel to a unique axis, the point-group rotation matrices can be partitioned into 2×2 and 1×1 diagonal blocks, the former corresponding to an operation of the plane group resulting from the projection of the space group down the unique axis. If, for example, the unique axis is parallel to **c**, we can decompose the scalar product in (1.4.2.19) and (1.4.2.20) as follows:

$$\begin{aligned} \mathbf{h}^T(\boldsymbol{P}_s\mathbf{r} + \mathbf{t}_s) = {}&\begin{pmatrix} h & k \end{pmatrix}\left[\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right] \\ &+ l(P_{33}z + t_3), \end{aligned} \tag{1.4.3.6}$$

where the first scalar product on the right-hand side of (1.4.3.6) contains the contribution of a plane group and the second product is the contribution of the unique axis itself. The above decomposition often leads to a convenient factorization of *A* and *B*, and is applicable to monoclinic, tetragonal and hexagonal families, the latter including rhombohedral space groups that are referred to hexagonal axes.

The symbols used in Tables A1.4.3.3, A1.4.3.5 and A1.4.3.6 are based on such decompositions. In those few cases where explicit expressions must be given we make use of the convention of replacing $\cos(2\pi u)$ by $\mathrm{c}(u)$ and $\sin(2\pi u)$ by $\mathrm{s}(u)$. For example, $\cos[2\pi(hy + kx)]$ *etc.* is given as $\mathrm{c}(hy + kx)$ *etc.* The symbols are defined below.

Monoclinic space groups (Table A1.4.3.3)

The following symbols are used in this system:

$$\begin{aligned} \mathrm{c}(hl) &= \cos[2\pi(hx + lz)], \quad & \mathrm{c}(hk) &= \cos[2\pi(hx + ky)] \\ \mathrm{s}(hl) &= \sin[2\pi(hx + lz)], \quad & \mathrm{s}(hk) &= \sin[2\pi(hx + ky)] \end{aligned} \tag{1.4.3.7}$$

so that any expression for *A* or *B* in the monoclinic system has the form $K\mathrm{p}(hl)\mathrm{q}(ky)$ or $K\mathrm{p}(hk)\mathrm{q}(lz)$ for the second or first setting, respectively, where p and q can each be a sine or a cosine and *K* is a numerical constant.

Tetragonal space groups (Table A1.4.3.5)

The most frequently occurring expressions in the summations for A and B in this system are of the form

$$\mathrm{P(pq)} = \mathrm{p}(2\pi hx)\mathrm{q}(2\pi ky) + \mathrm{p}(2\pi kx)\mathrm{q}(2\pi hy) \qquad (1.4.3.8)$$

and

$$\mathrm{M(pq)} = \mathrm{p}(2\pi hx)\mathrm{q}(2\pi ky) - \mathrm{p}(2\pi kx)\mathrm{q}(2\pi hy), \qquad (1.4.3.9)$$

where p and q can each be a sine or a cosine. These are typical contributions related to square plane groups.

Trigonal and hexagonal space groups (Table A1.4.3.6)

The contributions of plane hexagonal space groups to the first term in (1.4.3.6) are

$$\begin{aligned} p_1 &= hx + ky, \quad p_2 = kx + iy, \quad p_3 = ix + hy, \\ q_1 &= kx + hy, \quad q_2 = hx + iy, \quad q_3 = ix + ky, \end{aligned} \qquad (1.4.3.10)$$

where $i = -h - k$ (*IT* I, 1952). The symbols which represent the frequently occurring expressions in this family, and given in terms of (1.4.3.10), are

$$\begin{aligned} \mathrm{C}(hki) &= \cos(2\pi p_1) + \cos(2\pi p_2) + \cos(2\pi p_3) \\ \mathrm{C}(khi) &= \cos(2\pi q_1) + \cos(2\pi q_2) + \cos(2\pi q_3) \\ \mathrm{S}(hki) &= \sin(2\pi p_1) + \sin(2\pi p_2) + \sin(2\pi p_3) \\ \mathrm{S}(khi) &= \sin(2\pi q_1) + \sin(2\pi q_2) + \sin(2\pi q_3) \end{aligned} \qquad (1.4.3.11)$$

and these quite often appear as the following sums and differences:

$$\begin{aligned} \mathrm{PH(cc)} &= \mathrm{C}(hki) + \mathrm{C}(khi), \quad \mathrm{PH(ss)} = \mathrm{S}(hki) + \mathrm{S}(khi) \\ \mathrm{MH(cc)} &= \mathrm{C}(hki) - \mathrm{C}(khi), \quad \mathrm{MH(ss)} = \mathrm{S}(hki) - \mathrm{S}(khi). \end{aligned} \qquad (1.4.3.12)$$

The symbols defined in this section are briefly redefined in the appropriate tables, which also contain the conditions for vanishing symbols.

1.4.3.4. Arrangement of the tables

The expressions for A and B are usually presented in terms of the short symbols defined above for all the representations of the plane groups and space groups given in Volume A (*IT* A, 1983), and are fully consistent with the unit-cell choices and space-group origins employed in that volume. The tables are arranged by crystal families and the expressions appear in the order of the appearance of the corresponding plane and space groups in the space-group tables in *IT* A (1983).

The main items in a table entry, not necessarily in the following order, are: (i) the conventional space-group number, (ii) the short Hermann–Mauguin space-group symbol, (iii) brief remarks on the choice of the space-group origin and setting, where appropriate, (iv) the real (A) and imaginary (B) parts of the trigonometric structure factor, and (v) the parity of the hkl subset to which the expressions for A and B pertain. Full space-group symbols are given in the monoclinic system only, since they are indispensable for the recognition of the settings and glide planes appearing in the various representations of monoclinic space groups given in *IT* A (1983).

1.4.4. Symmetry in reciprocal space: space-group tables

1.4.4.1. Introduction

The purpose of this section, and the accompanying table, is to provide a representation of the 230 three-dimensional crystallographic space groups in terms of two fundamental quantities that characterize a weighted reciprocal lattice: (i) coordinates of point-symmetry-related points in the reciprocal lattice, and (ii) phase shifts of the weight functions that are associated with the translation parts of the various space-group operations. Table A1.4.4.1 in Appendix 1.4.4 collects the above information for all the space-group settings which are listed in *IT* A (1983) for the same choice of the space-group origins and following the same numbering scheme used in that volume. Table A1.4.4.1 was generated by computer using the space-group algorithm described by Shmueli (1984) and the space-group symbols given in Table A1.4.2.1 in Appendix 1.4.2. It is shown in a later part of this section that Table A1.4.4.1 can also be regarded as a table of symmetry groups in Fourier space, in the Bienenstock–Ewald (1962) sense which was mentioned in Section 1.4.1. The section is concluded with a brief description of the correspondence between Bravais-lattice types in direct and reciprocal spaces.

1.4.4.2. Arrangement of the space-group tables

Table A1.4.4.1 is subdivided into point-group sections and space-group subsections, as outlined below.

(i) *The point-group heading.* This heading contains a short Hermann–Mauguin symbol of a point group, the crystal system and the symbol of the Laue group with which the point group is associated. Each point-group heading is followed by the set of space groups which are isomorphic to the point group indicated, the set being enclosed within a box.

(ii) *The space-group heading.* This heading contains, for each space group listed in Volume A (*IT* A, 1983), the short Hermann–Mauguin symbol of the space group, its conventional space-group number and (in parentheses) the serial number of its representation in Volume A; this is also the serial number of the explicit space-group symbol in Table A1.4.2.1 from which the entry was derived. Additional items are full space-group symbols, given only for the monoclinic space groups in their settings that are given in Volume A (*IT*, 1983), and self-explanatory comments as required.

(iii) *The table entry.* In the context of the analysis in Section 1.4.2.2, the format of a table entry is: $\mathbf{h}^T\boldsymbol{P}_n : -\mathbf{h}^T\mathbf{t}_n$, where $(\boldsymbol{P}_n, \mathbf{t}_n)$ is the nth space-group operator, and the phase shift $\mathbf{h}^T\mathbf{t}_n$ is expressed in units of 2π [see equations (1.4.2.3) and (1.4.2.5)]. More explicitly, the general format of a table entry is

$$(n)\, h_n k_n l_n : -p_n q_n r_n/m. \qquad (1.4.4.1)$$

In (1.4.4.1), n is the serial number of the space-group operation to which this entry pertains and is the same as the number of the general Wyckoff position generated by this operation and given in *IT* A (1983) for the space group appearing in the space-group heading. The first part of an entry, $h_n k_n l_n$:, contains the coordinates of the reciprocal-lattice vector that was generated from the reference vector (hkl) by the rotation part of the nth space-group operation. These rotation parts of the table entries, for a given space group, thus constitute the set of reciprocal-lattice points that are generated by the corresponding point group (*not Laue group*). The second part of an entry is an abbreviation of the phase shift which is associated with the nth operation and thus

$$-p_n q_n r_n/m \quad \text{denotes} \quad -2\pi(hp_n + kq_n + lr_n)/m, \qquad (1.4.4.2)$$

where the fractions p_n/m, q_n/m and r_n/m are the components of the translation part $\mathbf{t}_n$ of the nth space-group operation. The

phase-shift part of an entry is given only if $(p_n q_n r_n)$ is *not* a vector in the direct lattice, since such a vector would give rise to a trivial phase shift (an integer multiple of 2π).

1.4.4.3. Effect of direct-space transformations

The phase shifts given in Table A1.4.4.1 depend on the translation parts of the space-group operations and these translations are determined, all or in part, by the choice of the space-group origin. It is a fairly easy matter to find the phase shifts that correspond to a given shift of the space-group origin in direct space, directly from Table A1.4.4.1. Moreover, it is also possible to modify the table entries so that a more general transformation, including a change of crystal axes as well as a shift of the space-group origin, can be directly accounted for. We employ here the frequently used concise notation due to Seitz (1935) (see also *IT* A, 1983).

Let the direct-space transformation be given by

$$\mathbf{r}_{\rm new} = \boldsymbol{T}\mathbf{r}_{\rm old} + \mathbf{v}, \tag{1.4.4.3}$$

where $\boldsymbol{T}$ is a nonsingular 3×3 matrix describing the change of the coordinate system and $\mathbf{v}$ is an origin-shift vector. The components of $\boldsymbol{T}$ and $\mathbf{v}$ are referred to the old system, and $\mathbf{r}_{\rm new}$ ($\mathbf{r}_{\rm old}$) is the position vector of a point in the crystal, referred to the new (old) system, respectively. If we denote a space-group operation referred to the new and old systems by $(\boldsymbol{P}_{\rm new}, \mathbf{t}_{\rm new})$ and $(\boldsymbol{P}_{\rm old}, \mathbf{t}_{\rm old})$, respectively, we have

$$(\boldsymbol{P}_{\rm new}, \mathbf{t}_{\rm new}) = (\boldsymbol{T}, \mathbf{v})(\boldsymbol{P}_{\rm old}, \mathbf{t}_{\rm old})(\boldsymbol{T}, \mathbf{v})^{-1} \tag{1.4.4.4}$$

$$= (\boldsymbol{T}\boldsymbol{P}_{\rm old}\boldsymbol{T}^{-1}, \mathbf{v} - \boldsymbol{T}\boldsymbol{P}_{\rm old}\boldsymbol{T}^{-1}\mathbf{v} + \boldsymbol{T}\mathbf{t}_{\rm old}). \tag{1.4.4.5}$$

It follows from (1.4.4.2) and (1.4.4.5) that if the old entry of Table A1.4.4.1 is given by

$$(n)\,\mathbf{h}^T\boldsymbol{P} : -\mathbf{h}^T\mathbf{t},$$

the transformed entry becomes

$$(n)\,\mathbf{h}^T\boldsymbol{T}\boldsymbol{P}\boldsymbol{T}^{-1} : \mathbf{h}^T\boldsymbol{T}\boldsymbol{P}\boldsymbol{T}^{-1}\mathbf{v} - \mathbf{h}^T\mathbf{v} - \mathbf{h}^T\boldsymbol{T}\mathbf{t}, \tag{1.4.4.6}$$

and in the important special cases of a pure change of setting ($\mathbf{v} = 0$) or a pure shift of the space-group origin ($\boldsymbol{T}$ is the unit matrix $\boldsymbol{I}$), (1.4.4.6) reduces to

$$(n)\,\mathbf{h}^T\boldsymbol{T}\boldsymbol{P}\boldsymbol{T}^{-1} : -\mathbf{h}^T\boldsymbol{T}\mathbf{t} \tag{1.4.4.7}$$

or

$$(n)\,\mathbf{h}^T\boldsymbol{P} : \mathbf{h}^T\boldsymbol{P}\mathbf{v} - \mathbf{h}^T\mathbf{v} - \mathbf{h}^T\mathbf{t}, \tag{1.4.4.8}$$

respectively. The rotation matrices $\boldsymbol{P}$ are readily obtained by visual or programmed inspection of the old entries: if, for example, $\mathbf{h}^T\boldsymbol{P}$ is *khl*, we must have $P_{21} = 1$, $P_{12} = 1$ and $P_{33} = 1$, the remaining P_{ij}'s being equal to zero. Similarly, if $\mathbf{h}^T\boldsymbol{P}$ is *kil*, where $i = -h - k$, we have

$$(kil) = (k, -h-k, l) = (hkl)\begin{pmatrix} 0 & \bar{1} & 0 \\ 1 & \bar{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The rotation matrices can also be obtained by reference to Part 7 and Tables 11.2.2.1 and 11.2.2.2 in Volume A (*IT* A, 2005).

As an example, consider the phase shifts corresponding to the operation No. (16) of the space group $P4/nmm$ (No. 129) in its two origins given in Volume A (*IT* A, 1983). For an Origin 2-to-Origin 1 transformation we find there $\mathbf{v} = (\frac{1}{4}, -\frac{1}{4}, 0)$ and the old Origin 2 entry in Table A1.4.4.1 is (16) *khl* ($\mathbf{t}$ is zero). The appropriate entry for the Origin 1 description of this operation should therefore be $\mathbf{h}^T\boldsymbol{P}\mathbf{v} - \mathbf{h}^T\mathbf{v} = k/4 - h/4 - h/4 + k/4 = -h/2 + k/2$, as given by (1.4.4.8), or $-(h+k)/2$ if a trivial shift of 2π is subtracted. The (new) Origin 1 entry thus becomes: (16) *khl*: $-110/2$, as listed in Table A1.4.4.1.

1.4.4.4. Symmetry in Fourier space

As shown below, Table A1.4.4.1 can also be regarded as a collection of the general equivalent positions of the symmetry groups of Fourier space, in the sense of the treatment by Bienenstock & Ewald (1962). This interpretation of the table is, however, restricted to the underlying periodic function being real and positive (see the latter reference). The symmetry formalism can be treated with the aid of the original 4×4 matrix notation, but it appears that a concise Seitz-type notation suits better the present introductory interpretation.

The symmetry dependence of the fundamental relationship (1.4.2.5)

$$\varphi(\mathbf{h}^T\boldsymbol{P}_n) = \varphi(\mathbf{h}) - 2\pi\mathbf{h}^T\mathbf{t}_n$$

is given by a table entry of the form: $(n)\,\mathbf{h}^T\boldsymbol{P} : -\mathbf{h}^T\mathbf{t}$, where the phase shift is given in units of 2π, and the structure-dependent phase $\varphi(\mathbf{h})$ is omitted. Defining a combination law analogous to Seitz's product of two operators of affine transformation:

$$[\mathbf{a}^T : b](\boldsymbol{R}, \mathbf{r}) = [\mathbf{a}^T\boldsymbol{R} : \mathbf{a}^T\mathbf{r} + b], \tag{1.4.4.9}$$

where $\boldsymbol{R}$ is a 3×3 matrix, $\mathbf{a}^T$ is a row vector, $\mathbf{r}$ is a column vector and b is a scalar, we can write the general form of a table entry as

$$[\mathbf{h}^T : \delta](\boldsymbol{P}, -\mathbf{t}) = [\mathbf{h}^T\boldsymbol{P} : -\mathbf{h}^T\mathbf{t} + \delta], \tag{1.4.4.10}$$

where δ is a constant phase shift which we take as zero. The positions $[\mathbf{h}^T : 0]$ and $[\mathbf{h}^T\boldsymbol{P} : -\mathbf{h}^T\mathbf{t}]$ are now related by the operation $(\boldsymbol{P}, -\mathbf{t})$ *via* the combination law (1.4.4.9), which is a shorthand transcription of the 4×4 matrix notation of Bienenstock & Ewald (1962), with the appropriate sign of $\mathbf{t}$.

Let us evaluate the result of a successive application of two such operators, say $(\boldsymbol{P}, -\mathbf{t})$ and $(\boldsymbol{Q}, -\mathbf{v})$ to the reference position $[\mathbf{h}^T : 0]$ in Fourier space:

$$[\mathbf{h}^T : 0](\boldsymbol{P}, -\mathbf{t})(\boldsymbol{Q}, -\mathbf{v}) = [\mathbf{h}^T : 0](\boldsymbol{P}\boldsymbol{Q}, -\boldsymbol{P}\mathbf{v} - \mathbf{t})$$

$$= [\mathbf{h}^T\boldsymbol{P}\boldsymbol{Q} : -\mathbf{h}^T\boldsymbol{P}\mathbf{v} - \mathbf{h}^T\mathbf{t}], \tag{1.4.4.11}$$

and perform an inverse operation:

$$[\mathbf{h}^T\boldsymbol{P} : -\mathbf{h}^T\mathbf{t}](\boldsymbol{P}, -\mathbf{t})^{-1} = [\mathbf{h}^T\boldsymbol{P} : -\mathbf{h}^T\mathbf{t}](\boldsymbol{P}^{-1}, \boldsymbol{P}^{-1}\mathbf{t})$$

$$= [\mathbf{h}^T\boldsymbol{P}\boldsymbol{P}^{-1} : \mathbf{h}^T\boldsymbol{P}\boldsymbol{P}^{-1}\mathbf{t} - \mathbf{h}^T\mathbf{t}]$$

$$= [\mathbf{h}^T : 0]. \tag{1.4.4.12}$$

These equations confirm the validity of the shorthand notation (1.4.4.9) and illustrate the group nature of the operators $(\boldsymbol{P}, -\mathbf{t})$ in the present context.

Table 1.4.4.1. *Correspondence between types of centring in direct and reciprocal lattices*

Direct lattice		Reciprocal lattice		
Lattice type(s)	Centring translations	Lattice type(s)	Restriction on *hkl*	Multiple unit cell
P, *R*		*P*, *R*		$\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$
A	$0, \frac{1}{2}, \frac{1}{2}$	*A*	$k+l=2n$	$\mathbf{a}^*$, $2\mathbf{b}^*$, $2\mathbf{c}^*$
B	$\frac{1}{2}, 0, \frac{1}{2}$	*B*	$h+l=2n$	$2\mathbf{a}^*$, $\mathbf{b}^*$, $2\mathbf{c}^*$
C	$\frac{1}{2}, \frac{1}{2}, 0$	*C*	$h+k=2n$	$2\mathbf{a}^*$, $2\mathbf{b}^*$, $\mathbf{c}^*$
I	$\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$	*F*	$h+k+l=2n$	$2\mathbf{a}^*$, $2\mathbf{b}^*$, $2\mathbf{c}^*$
F	$0, \frac{1}{2}, \frac{1}{2}$ $\frac{1}{2}, 0, \frac{1}{2}$ $\frac{1}{2}, \frac{1}{2}, 0$	*I*	$k+l=2n$ $h+l=2n$ $h+k=2n$	$2\mathbf{a}^*$, $2\mathbf{b}^*$, $2\mathbf{c}^*$
$R_{\rm hex}$	$\frac{2}{3}, \frac{1}{3}, \frac{1}{3}$ $\frac{1}{3}, \frac{2}{3}, \frac{2}{3}$	$R_{\rm hex}$	$-h+k+l=3n$	$3\mathbf{a}^*$, $3\mathbf{b}^*$, $3\mathbf{c}^*$

Following Bienenstock & Ewald, the operators $(\boldsymbol{P}, -\mathbf{t})$ are symmetry operators that act on the positions $[\mathbf{h}^T : 0]$ in Fourier space, provided they satisfy the following requirements: (i) the application of such an operator leaves the magnitude of the (generally) complex Fourier coefficient unchanged, and (ii) after g successive applications of an operator, where g is the order of its rotation part, the phase remains unchanged up to a shift by an integer multiple of 2π (a trivial phase shift, corresponding to a translation by a lattice vector in direct space).

If our function is the electron density in the crystal, the first requirement is obviously satisfied since $|F(\mathbf{h})| = |F(\mathbf{h}^T\boldsymbol{P})|$, where F is the structure factor [*cf.* equation (1.4.2.4)]. In order to make use of the second requirement in deriving permissible symmetry operators on Fourier space, all the relevant transformations, *i.e.* those which have rotation operators of the orders 1, 2, 3, 4 and 6, must be individually examined. A comprehensive example, covering most of the tetragonal system, can be found in Bienenstock & Ewald (1962).

It is of interest to illustrate the above process for a simple particular instance. Consider an operation, the rotation part of which involves a mirror plane, and assume that it is associated with the monoclinic system, in the second setting (unique axis b). We denote the operator by $(\boldsymbol{m}, -\mathbf{u})$, where $\mathbf{u}^T = (uvw)$, and the permissible values of u, v and w are to be determined. The operation is of order 2, and according to requirement (ii) above we have to evaluate

$$\begin{aligned}[\mathbf{h}^T : 0](\boldsymbol{m}, -\mathbf{u})^2 &= [\mathbf{h}^T : 0](\boldsymbol{I}, -\boldsymbol{m}\mathbf{u} - \mathbf{u}) \\ &= [\mathbf{h}^T : -\mathbf{h}^T(\boldsymbol{m} + \boldsymbol{I})\mathbf{u}] \\ &= [hkl : -2(hu + lw)], \end{aligned} \tag{1.4.4.13}$$

where

$$\boldsymbol{m} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \bar{1} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is the matrix representing the operation of reflection and $\boldsymbol{I}$ is the unit matrix. For $(\boldsymbol{m}, -\mathbf{u})$ to be an admissible symmetry operator, the phase-shift part of (1.4.4.13), *i.e.* $-2(hu + lw)$, must be an integer (multiple of 2π). The smallest non-negative values of u and w which satisfy this are the pairs: $u = w = 0$, $u = \frac{1}{2}$ and $w = 0$, $u = 0$ and $w = \frac{1}{2}$, and $u = w = \frac{1}{2}$. We have thus obtained four symmetry operators in Fourier space, which are identical (except for the sign of their translational parts) to those of the direct-space monoclinic mirror and glide-plane operations. The fact that the component v cancels out simply means that an arbitrary component of the phase shift can be added along the b^* axis; this is concurrent with arbitrary direct-space translations that appear in the characterization of individual types of space-group operations [see *e.g.* Koch & Fischer (2005)].

Each of the 230 space groups, which leaves invariant a (real and non-negative) function with the periodicity of the crystal, thus has its counterpart which determines the symmetry of the Fourier expansion coefficients of this function, with equivalent positions given in Table A1.4.4.1.

1.4.4.5. Relationships between direct and reciprocal Bravais lattices

Centred Bravais lattices in crystal space give rise to systematic absences of certain classes of reflections (*IT* I, 1952; *IT* A, 1983) and the corresponding points in the reciprocal lattice have accordingly zero weights. These absences are periodic in reciprocal space and their 'removal' from the reciprocal lattice results in a lattice which – like the direct one – must belong to one of the fourteen Bravais lattice types. This must be so since the point group of a crystal leaves its lattice – and also the associated reciprocal lattice – unchanged. The magnitudes of the structure factors (the weight functions) are also invariant under the operation of this point group.

The correspondence between the types of centring in direct and reciprocal lattices is given in Table 1.4.4.1.

Notes:

(i) The vectors $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$, appearing in the definition of the multiple unit cell in the reciprocal lattice, define this lattice *prior to the removal* of lattice points with zero weights (absences). All the restrictions on *hkl* pertain to indexing on $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$.

(ii) The centring type of the reciprocal lattice refers to the multiple unit cell given in the table.

(iii) The centring type denoted by $R_{\rm hex}$ is a representation of the rhombohedral lattice R by a triple hexagonal unit cell, in the obverse setting (*IT* I, 1952), *i.e.* according to the transformation

$$\begin{aligned} \mathbf{a} &= \mathbf{a}_R - \mathbf{b}_R \\ \mathbf{b} &= \mathbf{b}_R - \mathbf{c}_R \\ \mathbf{c} &= \mathbf{a}_R + \mathbf{b}_R + \mathbf{c}_R, \end{aligned} \tag{1.4.4.14}$$

where $\mathbf{a}_R$, $\mathbf{b}_R$ and $\mathbf{c}_R$ pertain to a primitive unit cell in the rhombohedral lattice R.

The corresponding multiple reciprocal cell, with centring denoted by $R_{\rm hex}$, contains nine lattice points with coordinates 000, 021, 012, 101, 202, 110, 220, 211 and 122 – indexed on the usual reciprocal to the triple hexagonal unit cell defined by (1.4.4.14). Detailed derivations of these correspondences are given by Buerger (1942), and an elementary proof of the reciprocity of I and F lattices can be found, *e.g.*, in pamphlet No. 4 of the Commission on Crystallographic Teaching (Authier, 1981).

Intuitive proofs follow directly from the restrictions on *hkl*, given in Table 1.4.4.1.

APPENDIX A1.4.1
Comments on the preparation and usage of the tables

By U. Shmueli

The straightforward but rather extensive calculations and text processing related to Tables A1.4.3.1 through A1.4.3.7 and Table A1.4.4.1 in Appendices 1.4.3 and 1.4.4, respectively, were performed with the aid of a combination of FORTRAN and REDUCE (Hearn, 1973) programs, designed so as to enable the author to produce the table entries directly from a space-group symbol and with a minimum amount of intermediate manual intervention. The first stage of the calculation, the generation of a space group (coordinates of the equivalent positions), was accomplished with the program *SPGRGEN*, the algorithm of which was described in some detail elsewhere (Shmueli, 1984). A complete list of computer-adapted space-group symbols, processed by *SPGRGEN* and not given in the latter reference, is presented in Table A1.4.2.1 of Appendix 1.4.2.

The generation of the space group is followed by a construction of symbolic expressions for the scalar products $\mathbf{h}^T(\boldsymbol{P}\mathbf{r}+\mathbf{t})$; *e.g.* for position No. (13) in the space group $P4_132$ (No. 213, *IT* I, 1952, *IT* A, 1983), this scalar product is given by $h(\frac{3}{4}+y)+k(\frac{1}{4}+x)+l(\frac{1}{4}-z)$. The construction of the various table entries consists of expanding the sines and cosines of these scalar products, performing the required summations, and simplifying the result where possible. The construction of the scalar products in a FORTRAN program is fairly easy and the extremely tedious trigonometric calculations required by equations (1.4.2.19) and (1.4.2.20) can be readily performed with the aid of one of several available computer-algebraic languages (for a review, see *Computers in the New Laboratory – a Nature Survey*, 1981); the REDUCE language was employed for the above purpose.

Since the REDUCE programs required for the summations in (1.4.2.19) and (1.4.2.20) for the various space groups were seen to have much in common, it was decided to construct a FORTRAN interface which would process the space-group input and prepare automatically REDUCE programs for the algebraic work. The least straightforward problem encountered during this work was the need to 'convince' the interface to generate *hkl* parity assignments which are appropriate to the space-group information input. This was solved for all the crystal families except the hexagonal by setting up a 'basis' of the form: $h/2, k/2, l/2, (k+l)/2, \ldots, (h+k+l)/4$ and representing the translation parts of the scalar products, $\mathbf{h}^T\mathbf{t}$, as sums of such 'basis functions'. A subsequent construction of an automatic parity routine proved to be easy and the interface could thus produce any number of REDUCE programs for the summations in (1.4.2.19) and (1.4.2.20) using a list of space-group symbols as the sole input. These included trigonal and hexagonal space groups with translation components of $\frac{1}{2}$. This approach seemed to be too awkward for some space groups containing threefold and sixfold screw axes, and these were treated individually.

There is little to say about the REDUCE programs, except that the output they generate is at the same level of trigonometric complexity as the expressions for A and B appearing in Volume I (*IT* I, 1952). This could have been improved by making use of the pattern-matching capabilities that are incorporated in REDUCE, but it was found more convenient to construct a FORTRAN interpreter which would detect in the REDUCE output the basic building blocks of the trigonometric structure factors (see Section 1.4.3.3) and perform the required transformations.

Tables A1.4.3.1–A1.4.3.7 were thus constructed with the aid of a chain composed of (i) a space-group generating routine, (ii) a FORTRAN interface, which processes the space-group input and 'writes' a complete REDUCE program, (iii) execution of the REDUCE program and (iv) a FORTRAN interpreter of the REDUCE output in terms of the abbreviated symbols to be used in the tables. The computation was at a 'one-group-at-a-time' basis and the automation of its repetition was performed by means of procedural constructs at the operating-system level. The construction of Table A1.4.4.1 involved only the preliminary stage of the processing of the space-group information by the FORTRAN interface. All the computations were carried out on a Cyber 170–855 at the Tel Aviv University Computation Center.

It is of some importance to comment on the recommended usage of the tables included in this chapter in automatic computations. If, for example, we wish to compute the expression: $A=-8(\mathrm{Escs}+\mathrm{Ossc})$, use can be made of the facility provided by most versions of FORTRAN of transferring subprogram names as parameters of a FUNCTION. We thus need only two FUNCTIONs for any calculation of A and B for a cubic space group, one FUNCTION for the block of even permutations of x, y and z:

```
      FUNCTION E(P,Q,R)
      EXTERNAL SIN,COS
      COMMON/TSF/TPH,TPK,TPL,X,Y,Z
      E = P(TPH * X) * Q(TPK * Y) * R(TPL * Z)
     1 + P(TPH * Z) * Q(TPK * X) * R(TPL * Y)
     2 + P(TPH * Y) * Q(TPK * Z) * R(TPL * X)
      RETURN
      END
```

where TPH, TPK and TPL denote $2\pi h$, $2\pi k$ and $2\pi l$, respectively, and a similar FUNCTION, say O(P,Q,R), for the block of odd permutations of x, y and z. The calling statement in the calling (sub)program can thus be:

```
A = -8 * (E(SIN,COS,SIN) + O(SIN,SIN,COS)).
```

A small number of such FUNCTIONs suffices for all the space-group-specific computations that involve trigonometric structure factors.

APPENDIX A1.4.2
Space-group symbols for numeric and symbolic computations

A1.4.2.1. Introduction

By U. Shmueli, S. R. Hall and R. W. Grosse-Kunstleve

This appendix lists two sets of computer-adapted space-group symbols which are implemented in existing crystallographic software and can be employed in the automated generation of space-group representations. The computer generation of space-group symmetry information is of well known importance in many crystallographic calculations, numeric as well as symbolic. A prerequisite for a computer program that generates this information is a set of computer-adapted space-group symbols which are based on the generating elements of the space group to be derived. The sets of symbols to be presented are:

(i) *Explicit symbols*. These symbols are based on the classification of crystallographic point groups and space groups by Zachariasen (1945). These symbols are termed *explicit* because they contain in an explicit manner the rotation and translation parts of the space-group generators of the space group to be derived and used. These computer-adapted explicit symbols were proposed by Shmueli (1984), who also describes in detail their implementation in the program *SPGRGEN*. This program was used for the automatic preparation of the structure-factor tables for the 17 plane groups and 230 space groups, listed in Appendix 1.4.3, and the 230 space groups in reciprocal space, listed in Appendix 1.4.4. The explicit symbols presented in this appendix are adapted to the 306 representations of the 230 space groups as presented in *IT* A (1983) with regard to the standard settings and choice of space-group origins.

The symmetry-generating algorithm underlying the explicit symbols, and their definition, are given in Section A1.4.2.2 of this appendix and the explicit symbols are listed in Table A1.4.2.1.

(ii) *Hall symbols*. These symbols are based on the implied-origin notation of Hall (1981*a*,*b*), who also describes in detail the algorithm implemented in the program *SGNAME* (Hall, 1981*a*). In the first edition of *IT* B (1993), the term 'concise space-group symbols' was used for this notation. In recent years, however, the term 'Hall symbols' has come into use in symmetry papers (Altermatt & Brown, 1987; Grosse-Kunstleve, 1999), software applications (Hovmöller, 1992; Grosse-Kunstleve, 1995; Larine *et al.*, 1995; Dowty, 1997) and data-handling approaches (Bourne *et al.*, 1998). This term has therefore been adopted for the second edition.

The main difference in the definition of the Hall symbols between this edition and the first edition of *IT* B is the generalization of the origin-shift vector to a full change-of-basis matrix. The examples have been expanded to show how this matrix is applied. The notation has also been made more consistent, and a typographical error in a default axis direction has been corrected.[1] The lattice centring symbol 'H' has been added to Table A1.4.2.2. In addition, Hall symbols are now provided for 530 settings to include all settings from Table 4.3.1 of *IT* A (1983). Namely, all nonstandard symbols for the monoclinic and orthorhombic space groups are included.

Some of the space-group symbols listed in Table A1.4.2.7 differ from those listed in Table B.6 (p. 119) of the first edition of *IT* B. This is because the symmetry of many space groups can be represented by more than one subset of 'generator' elements and these lead to different Hall symbols. The symbols listed in this edition have been selected after first sorting the symmetry elements into a strictly prescribed order based on the shape of their Seitz matrices, whereas those in Table B.6 were selected from symmetry elements in the order of *IT* I (1965). Software for selecting the Hall symbols listed in Table A1.4.2.7 is freely available (Hall, 1997). These symbols and their equivalents in the first edition of *IT* B will generate identical symmetry elements, but the former may be used as a reference table in a strict mapping procedure between different symmetry representations (Hall *et al.*, 2000).

The Hall symbols are defined in Section A1.4.2.3 of this appendix and are listed in Table A1.4.2.7.

A1.4.2.2. Explicit symbols

By U. Shmueli

As shown elsewhere (Shmueli, 1984), the set of representative operators of a crystallographic space group [*i.e.* the set that is listed for each space group in the symmetry tables of *IT* A (1983) and automatically regenerated for the purpose of compiling the symmetry tables in the present chapter] may have one of the following forms:

$$\begin{aligned} &\{(\mathbf{Q}, \mathbf{u})\}, \\ &\{(\mathbf{Q}, \mathbf{u})\} \times \{(\mathbf{R}, \mathbf{v})\}, \quad \text{or} \\ &\{(\mathbf{P}, \mathbf{t})\} \times [\{(\mathbf{Q}, \mathbf{u})\} \times \{(\mathbf{R}, \mathbf{v})\}], \end{aligned} \tag{A1.4.2.1}$$

where **P**, **Q** and **R** are point-group operators, and **t**, **u** and **v** are zero vectors or translations not belonging to the lattice-translations subgroup. Each of the forms in (A1.4.2.1), enclosed in braces, is evaluated as, *e.g.*,

$$\{(\mathbf{P}, \mathbf{t})\} = \{(\mathbf{I}, \mathbf{0}), (\mathbf{P}, \mathbf{t}), (\mathbf{P}, \mathbf{t})^2, \ldots, (\mathbf{P}, \mathbf{t})^{g-1}\}, \tag{A1.4.2.2}$$

where **I** is a unit operator and g is the order of the rotation operator **P** (*i.e.* $\mathbf{P}^g = \mathbf{I}$). The representative operations of the space group are evaluated by expanding the generators into cyclic groups, as in (A1.4.2.2), and forming, as needed, ordered products of the expanded groups as indicated in (A1.4.2.1) and explained in detail in the original article (Shmueli, 1984). The rotation and translation parts of the generators (**P**, **t**), (**Q**, **u**) and (**R**, **v**) presented here were adapted to the settings and choices of origin used in the main symmetry tables of *IT* A (1983).

The general structure of a three-generator symbol, corresponding to the last line of (A1.4.2.1), as represented in Table A1.4.2.1, is

$$\mathrm{LSC\$r_1Pt_1t_2t_3\$r_2Qu_1u_2u_3\$r_3Rv_1v_2v_3}, \tag{A1.4.2.3}$$

where

L – lattice type; can be P, A, B, C, I, F, or R. The symbol R is used only for the seven rhombohedral space groups in their representations in rhombohedral and hexagonal axes [obverse setting (*IT* I, 1952)].

S – crystal system; can be A (triclinic), M (monoclinic), O (orthorhombic), T (tetragonal), R (trigonal), H (hexagonal) or C (cubic).

C – status of centrosymmetry; can be C or N according as the space group is centrosymmetric or noncentrosymmetric, respectively.

$ – this character is followed by six characters that define a generator of the space group.

r_i – indicator of the type of rotation that follows: r_i is P or I according as the rotation part of the ith generator is proper or improper, respectively.

P, Q, R – two-character symbols of matrix representations of the point-group rotation operators **P**, **Q** and **R**, respectively (see below).

$t_1t_2t_3$, $u_1u_2u_3$, $v_1v_2v_3$ – components of the translation parts of the generators, given in units of $\frac{1}{12}$; *e.g.* the translation part $(0\ \frac{1}{2}\ \frac{3}{4})$ is given in Table A1.4.2.1 as 069. *An exception*: $(0\ 0\ \frac{5}{6})$ is denoted by 005 and not by 0010.

The two-character symbols for the matrices of rotation, which appear in the explicit space-group symbols in Table A1.4.2.1, are defined as follows:

[1] The correct default axis direction $\mathbf{a} - \mathbf{b}$ of an N preceded by 3 or 6 replaces $\mathbf{a} + \mathbf{b}$ on p. 117, right-hand column, line 4, in the first edition of *IT* B.

Table A1.4.2.1. *Explicit symbols*

No.	Short Hermann–Mauguin symbol	Comments	Explicit symbols
1	$P1$		PAN$P1A000
2	$P\bar{1}$		PAC$I1A000
3	$P2$	$P121$	PMN$P2B000
3	$P2$	$P112$	PMN$P2C000
4	$P2_1$	$P12_11$	PMN$P2B060
4	$P2_1$	$P112_1$	PMN$P2C006
5	$C2$	$C121$	CMN$P2B000
5	$C2$	$A121$	AMN$P2B000
5	$C2$	$I121$	IMN$P2B000
5	$C2$	$A112$	AMN$P2C000
5	$C2$	$B112$	BMN$P2C000
5	$C2$	$I112$	IMN$P2C000
6	Pm	$P1m1$	PMN$I2B000
6	Pm	$P11m$	PMN$I2C000
7	Pc	$P1c1$	PMN$I2B006
7	Pc	$P1n1$	PMN$I2B606
7	Pc	$P1a1$	PMN$I2B600
7	Pc	$P11a$	PMN$I2C600
7	Pc	$P11n$	PMN$I2C660
7	Pc	$P11b$	PMN$I2C060
8	Cm	$C1m1$	CMN$I2B000
8	Cm	$A1m1$	AMN$I2B000
8	Cm	$I1m1$	IMN$I2B000
8	Cm	$A11m$	AMN$I2C000
8	Cm	$B11m$	BMN$I2C000
8	Cm	$I11m$	IMN$I2C000
9	Cc	$C1c1$	CMN$I2B006
9	Cc	$A1n1$	AMN$I2B606
9	Cc	$I1a1$	IMN$I2B600
9	Cc	$A11a$	AMN$I2C600
9	Cc	$B11n$	BMN$I2C660
9	Cc	$I11b$	IMN$I2C060
10	$P2/m$	$P12/m1$	PMC$I1A000$P2B000
10	$P2/m$	$P112/m$	PMC$I1A000$P2C000
11	$P2_1/m$	$P12_1/m1$	PMC$I1A000$P2B060
11	$P2_1/m$	$P112_1/m$	PMC$I1A000$P2C006
12	$C2/m$	$C12/m1$	CMC$I1A000$P2B000
12	$C2/m$	$A12/m1$	AMC$I1A000$P2B000
12	$C2/m$	$I12/m1$	IMC$I1A000$P2B000
12	$C2/m$	$A112/m$	AMC$I1A000$P2C000
12	$C2/m$	$B112/m$	BMC$I1A000$P2C000
12	$C2/m$	$I112/m$	IMC$I1A000$P2C000
13	$P2/c$	$P12/c1$	PMC$I1A000$P2B006
13	$P2/c$	$P12/n1$	PMC$I1A000$P2B606
13	$P2/c$	$P12/a1$	PMC$I1A000$P2B600
13	$P2/c$	$P112/a$	PMC$I1A000$P2C600
13	$P2/c$	$P112/n$	PMC$I1A000$P2C660
13	$P2/c$	$P112/b$	PMC$I1A000$P2C060
14	$P2_1/c$	$P12_1/c1$	PMC$I1A000$P2B066
14	$P2_1/c$	$P12_1/n1$	PMC$I1A000$P2B666
14	$P2_1/c$	$P12_1/a1$	PMC$I1A000$P2B660
14	$P2_1/c$	$P112_1/a$	PMC$I1A000$P2C606
14	$P2_1/c$	$P112_1/n$	PMC$I1A000$P2C666
14	$P2_1/c$	$P112_1/b$	PMC$I1A000$P2C066
15	$C2/c$	$C12/c1$	CMC$I1A000$P2B006
15	$C2/c$	$A12/n1$	AMC$I1A000$P2B606
15	$C2/c$	$I12/a1$	IMC$I1A000$P2B600
15	$C2/c$	$A112/a$	AMC$I1A000$P2C600
15	$C2/c$	$B112/n$	BMC$I1A000$P2C660
15	$C2/c$	$I112/b$	IMC$I1A000$P2C060
16	$P222$		PON$P2C000$P2A000
17	$P222_1$		PON$P2C006$P2A000
18	$P2_12_12$		PON$P2C000$P2A660
19	$P2_12_12_1$		PON$P2C606$P2A660
20	$C222_1$		CON$P2C006$P2A000
21	$C222$		CON$P2C000$P2A000
22	$F222$		FON$P2C000$P2A000
23	$I222$		ION$P2C000$P2A000
24	$I2_12_12_1$		ION$P2C606$P2A660
25	$Pmm2$		PON$P2C000$I2A000
26	$Pmc2_1$		PON$P2C006$I2A000
27	$Pcc2$		PON$P2C000$I2A006
28	$Pma2$		PON$P2C000$I2A600
29	$Pca2_1$		PON$P2C006$I2A606
30	$Pnc2$		PON$P2C000$I2A066
31	$Pmn2_1$		PON$P2C606$I2A000
32	$Pba2$		PON$P2C000$I2A660
33	$Pna2_1$		PON$P2C006$I2A666
34	$Pnn2$		PON$P2C000$I2A666
35	$Cmm2$		CON$P2C000$I2A000
36	$Cmc2_1$		CON$P2C006$I2A000
37	$Ccc2$		CON$P2C000$I2A006
38	$Amm2$		AON$P2C000$I2A000
39	$Abm2$		AON$P2C000$I2A060
40	$Ama2$		AON$P2C000$I2A600
41	$Aba2$		AON$P2C000$I2A660
42	$Fmm2$		FON$P2C000$I2A000
43	$Fdd2$		FON$P2C000$I2A333
44	$Imm2$		ION$P2C000$I2A000
45	$Iba2$		ION$P2C000$I2A660
46	$Ima2$		ION$P2C000$I2A600
47	$Pmmm$		POC$I1A000$P2C000$P2A000
48	$Pnnn$	Origin 1	POC$I1A666$P2C000$P2A000
48	$Pnnn$	Origin 2	POC$I1A000$P2C660$P2A066
49	$Pccm$		POC$I1A000$P2C000$P2A006
50	$Pban$	Origin 1	POC$I1A660$P2C000$P2A000
50	$Pban$	Origin 2	POC$I1A000$P2C660$P2A060
51	$Pmma$		POC$I1A000$P2C600$P2A600
52	$Pnna$		POC$I1A000$P2C600$P2A066
53	$Pmna$		POC$I1A000$P2C606$P2A000
54	$Pcca$		POC$I1A000$P2C600$P2A606
55	$Pbam$		POC$I1A000$P2C000$P2A660
56	$Pccn$		POC$I1A000$P2C660$P2A606
57	$Pbcm$		POC$I1A000$P2C006$P2A060
58	$Pnnm$		POC$I1A000$P2C000$P2A666
59	$Pmmn$	Origin 1	POC$I1A660$P2C000$P2A660
59	$Pmmn$	Origin 2	POC$I1A000$P2C660$P2A600
60	$Pbcn$		POC$I1A000$P2C666$P2A660
61	$Pbca$		POC$I1A000$P2C606$P2A660
62	$Pnma$		POC$I1A000$P2C606$P2A666
63	$Cmcm$		COC$I1A000$P2C006$P2A000
64	$Cmca$		COC$I1A000$P2C066$P2A000
65	$Cmmm$		COC$I1A000$P2C000$P2A000
66	$Cccm$		COC$I1A000$P2C000$P2A006
67	$Cmma$		COC$I1A000$P2C060$P2A000
68	$Ccca$	Origin 1	COC$I1A066$P2C660$P2A660

Table A1.4.2.1 (*cont.*)

No.	Short Hermann–Mauguin symbol	Comments	Explicit symbols
68	*Ccca*	Origin 2	COC$I1A000$P2C600$P2A606
69	*Fmmm*		FOC$I1A000$P2C000$P2A000
70	*Fddd*	Origin 1	FOC$I1A333$P2C000$P2A000
70	*Fddd*	Origin 2	FOC$I1A000$P2C990$P2A099
71	*Immm*		IOC$I1A000$P2C000$P2A000
72	*Ibam*		IOC$I1A000$P2C000$P2A660
73	*Ibca*		IOC$I1A000$P2C606$P2A660
74	*Imma*		IOC$I1A000$P2C060$P2A000
75	$P4$		PTN$P4C000
76	$P4_1$		PTN$P4C003
77	$P4_2$		PTN$P4C006
78	$P4_3$		PTN$P4C009
79	$I4$		ITN$P4C000
80	$I4_1$		ITN$P4C063
81	$P\bar{4}$		PTN$I4C000
82	$I\bar{4}$		ITN$I4C000
83	$P4/m$		PTC$I1A000$P4C000
84	$P4_2/m$		PTC$I1A000$P4C006
85	$P4/n$	Origin 1	PTC$I1A660$P4C660
85	$P4/n$	Origin 2	PTC$I1A000$P4C600
86	$P4_2/n$	Origin 1	PTC$I1A666$P4C666
86	$P4_2/n$	Origin 2	PTC$I1A000$P4C066
87	$I4/m$		ITC$I1A000$P4C000
88	$I4_1/a$	Origin 1	ITC$I1A063$P4C063
88	$I4_1/a$	Origin 2	ITC$I1A000$P4C933
89	$P422$		PTN$P4C000$P2A000
90	$P42_12$		PTN$P4C660$P2A660
91	$P4_122$		PTN$P4C003$P2A006
92	$P4_12_12$		PTN$P4C663$P2A669
93	$P4_222$		PTN$P4C006$P2A000
94	$P4_22_12$		PTN$P4C666$P2A666
95	$P4_322$		PTN$P4C009$P2A006
96	$P4_32_12$		PTN$P4C669$P2A663
97	$I422$		ITN$P4C000$P2A000
98	$I4_122$		ITN$P4C063$P2A063
99	$P4mm$		PTN$P4C000$I2A000
100	$P4bm$		PTN$P4C000$I2A660
101	$P4_2cm$		PTN$P4C006$I2A006
102	$P4_2nm$		PTN$P4C666$I2A666
103	$P4cc$		PTN$P4C000$I2A006
104	$P4nc$		PTN$P4C000$I2A666
105	$P4_2mc$		PTN$P4C006$I2A000
106	$P4_2bc$		PTN$P4C006$I2A660
107	$I4mm$		ITN$P4C000$I2A000
108	$I4cm$		ITN$P4C000$I2A006
109	$I4_1md$		ITN$P4C063$I2A666
110	$I4_1cd$		ITN$P4C063$I2A660
111	$P\bar{4}2m$		PTN$I4C000$P2A000
112	$P\bar{4}2c$		PTN$I4C000$P2A006
113	$P\bar{4}2_1m$		PTN$I4C000$P2A660
114	$P\bar{4}2_1c$		PTN$I4C000$P2A666
115	$P\bar{4}m2$		PTN$I4C000$P2D000
116	$P\bar{4}c2$		PTN$I4C000$P2D006
117	$P\bar{4}b2$		PTN$I4C000$P2D660
118	$P\bar{4}n2$		PTN$I4C000$P2D666
119	$I\bar{4}m2$		ITN$I4C000$P2D000
120	$I\bar{4}c2$		ITN$I4C000$P2D006
121	$I\bar{4}2m$		ITN$I4C000$P2A000
122	$I\bar{4}2d$		ITN$I4C000$P2A609
123	$P4/mmm$		PTC$I1A000$P4C000$P2A000
124	$P4/mcc$		PTC$I1A000$P4C000$P2A006
125	$P4/nbm$	Origin 1	PTC$I1A660$P4C000$P2A000
125	$P4/nbm$	Origin 2	PTC$I1A000$P4C600$P2A060
126	$P4/nnc$	Origin 1	PTC$I1A666$P4C000$P2A000
126	$P4/nnc$	Origin 2	PTC$I1A000$P4C600$P2A066
127	$P4/mbm$		PTC$I1A000$P4C000$P2A660
128	$P4/mnc$		PTC$I1A000$P4C000$P2A666
129	$P4/nmm$	Origin 1	PTC$I1A660$P4C660$P2A660
129	$P4/nmm$	Origin 2	PTC$I1A000$P4C600$P2A600
130	$P4/ncc$	Origin 1	PTC$I1A660$P4C660$P2A666
130	$P4/ncc$	Origin 2	PTC$I1A000$P4C600$P2A606
131	$P4_2/mmc$		PTC$I1A000$P4C006$P2A000
132	$P4_2/mcm$		PTC$I1A000$P4C006$P2A006
133	$P4_2/nbc$	Origin 1	PTC$I1A666$P4C666$P2A006
133	$P4_2/nbc$	Origin 2	PTC$I1A000$P4C606$P2A060
134	$P4_2/nnm$	Origin 1	PTC$I1A666$P4C666$P2A000
134	$P4_2/nnm$	Origin 2	PTC$I1A000$P4C606$P2A066
135	$P4_2/mbc$		PTC$I1A000$P4C006$P2A660
136	$P4_2/mnm$		PTC$I1A000$P4C666$P2A666
137	$P4_2/nmc$	Origin 1	PTC$I1A666$P4C666$P2A666
137	$P4_2/nmc$	Origin 2	PTC$I1A000$P4C606$P2A600
138	$P4_2/ncm$	Origin 1	PTC$I1A666$P4C666$P2A660
138	$P4_2/ncm$	Origin 2	PTC$I1A000$P4C606$P2A606
139	$I4/mmm$		ITC$I1A000$P4C000$P2A000
140	$I4/mcm$		ITC$I1A000$P4C000$P2A006
141	$I4_1/amd$	Origin 1	ITC$I1A063$P4C063$P2A063
141	$I4_1/amd$	Origin 2	ITC$I1A000$P4C393$P2A000
142	$I4_1/acd$	Origin 1	ITC$I1A063$P4C063$P2A069
142	$I4_1/acd$	Origin 2	ITC$I1A000$P4C393$P2A006
143	$P3$		PRN$P3C000
144	$P3_1$		PRN$P3C004
145	$P3_2$		PRN$P3C008
146	$R3$	Hexagonal axes	RRN$P3C000
146	$R3$	Rhombohedral axes	PRN$P3Q000
147	$P\bar{3}$		PRC$I3C000
148	$R\bar{3}$	Hexagonal axes	RRC$I3C000
148	$R\bar{3}$	Rhombohedral axes	PRC$I3Q000
149	$P312$		PRN$P3C000$P2G000
150	$P321$		PRN$P3C000$P2F000
151	$P3_112$		PRN$P3C004$P2G000
152	$P3_121$		PRN$P3C004$P2F008
153	$P3_212$		PRN$P3C008$P2G000
154	$P3_221$		PRN$P3C008$P2F004
155	$R32$	Hexagonal axes	RRN$P3C000$P2F000
155	$R32$	Rhombohedral axes	PRN$P3Q000$P2E000
156	$P3m1$		PRN$P3C000$I2F000
157	$P31m$		PRN$P3C000$I2G000
158	$P3c1$		PRN$P3C000$I2F006
159	$P31c$		PRN$P3C000$I2G006
160	$R3m$	Hexagonal axes	RRN$P3C000$I2F000
160	$R3m$	Rhombohedral axes	PRN$P3Q000$I2E000
161	$R3c$	Hexagonal axes	RRN$P3C000$I2F006
161	$R3c$	Rhombohedral axes	PRN$P3Q000$I2E666
162	$P\bar{3}1m$		PRC$I3C000$P2G000
163	$P\bar{3}1c$		PRC$I3C000$P2G006
164	$P\bar{3}m1$		PRC$I3C000$P2F000

Table A1.4.2.1 (*cont.*)

No.	Short Hermann–Mauguin symbol	Comments	Explicit symbols
165	$P\bar{3}c1$		PRC$I3C000$P2F006
166	$R\bar{3}m$	Hexagonal axes	RRC$I3C000$P2F000
166	$R\bar{3}m$	Rhombohedral axes	PRC$I3Q000$P2E000
167	$R\bar{3}c$	Hexagonal axes	RRC$I3C000$P2F006
167	$R\bar{3}c$	Rhombohedral axes	PRC$I3Q000$P2E666
168	$P6$		PHN$P6C000
169	$P6_1$		PHN$P6C002
170	$P6_5$		PHN$P6C005
171	$P6_2$		PHN$P6C004
172	$P6_4$		PHN$P6C008
173	$P6_3$		PHN$P6C006
174	$P\bar{6}$		PHN$I6C000
175	$P6/m$		PHC$I1A000$P6C000
176	$P6_3/m$		PHC$I1A000$P6C006
177	$P622$		PHN$P6C000$P2F000
178	$P6_122$		PHN$P6C002$P2F000
179	$P6_522$		PHN$P6C005$P2F000
180	$P6_222$		PHN$P6C004$P2F000
181	$P6_422$		PHN$P6C008$P2F000
182	$P6_322$		PHN$P6C006$P2F000
183	$P6mm$		PHN$P6C000$I2F000
184	$P6cc$		PHN$P6C000$I2F006
185	$P6_3cm$		PHN$P6C006$I2F006
186	$P6_3mc$		PHN$P6C006$I2F000
187	$P\bar{6}m2$		PHN$I6C000$P2G000
188	$P\bar{6}c2$		PHN$I6C006$P2G000
189	$P\bar{6}2m$		PHN$I6C000$P2F000
190	$P\bar{6}2c$		PHN$I6C006$P2F000
191	$P6/mmm$		PHC$I1A000$P6C000$P2F000
192	$P6/mcc$		PHC$I1A000$P6C000$P2F006
193	$P6_3/mcm$		PHC$I1A000$P6C006$P2F006
194	$P6_3/mmc$		PHC$I1A000$P6C006$P2F000
195	$P23$		PCN$P3Q000$P2C000$P2A000
196	$F23$		FCN$P3Q000$P2C000$P2A000
197	$I23$		ICN$P3Q000$P2C000$P2A000
198	$P2_13$		PCN$P3Q000$P2C606$P2A660
199	$I2_13$		ICN$P3Q000$P2C606$P2A660
200	$Pm\bar{3}$		PCC$I3Q000$P2C000$P2A000
201	$Pn\bar{3}$	Origin 1	PCC$I3Q666$P2C000$P2A000
201	$Pn\bar{3}$	Origin 2	PCC$I3Q000$P2C660$P2A066
202	$Fm\bar{3}$		FCC$I3Q000$P2C000$P2A000
203	$Fd\bar{3}$	Origin 1	FCC$I3Q333$P2C000$P2A000
203	$Fd\bar{3}$	Origin 2	FCC$I3Q000$P2C330$P2A033
204	$Im\bar{3}$		ICC$I3Q000$P2C000$P2A000
205	$Pa\bar{3}$		PCC$I3Q000$P2C606$P2A660
206	$Ia\bar{3}$		ICC$I3Q000$P2C606$P2A660
207	$P432$		PCN$P3Q000$P4C000$P2D000
208	$P4_232$		PCN$P3Q000$P4C666$P2D666
209	$F432$		FCN$P3Q000$P4C000$P2D000
210	$F4_132$		FCN$P3Q000$P4C993$P2D939
211	$I432$		ICN$P3Q000$P4C000$P2D000
212	$P4_332$		PCN$P3Q000$P4C939$P2D399
213	$P4_132$		PCN$P3Q000$P4C393$P2D933
214	$I4_132$		ICN$P3Q000$P4C393$P2D933
215	$P\bar{4}3m$		PCN$P3Q000$I4C000$I2D000
216	$F\bar{4}3m$		FCN$P3Q000$I4C000$I2D000
217	$I\bar{4}3m$		ICN$P3Q000$I4C000$I2D000
218	$P\bar{4}3n$		PCN$P3Q000$I4C666$I2D666
219	$F\bar{4}3c$		FCN$P3Q000$I4C666$I2D666
220	$I\bar{4}3d$		ICN$P3Q000$I4C939$I2D399
221	$Pm\bar{3}m$		PCC$I3Q000$P4C000$P2D000
222	$Pn\bar{3}n$	Origin 1	PCC$I3Q666$P4C000$P2D000
222	$Pn\bar{3}n$	Origin 2	PCC$I3Q000$P4C600$P2D006
223	$Pm\bar{3}n$		PCC$I3Q000$P4C666$P2D666
224	$Pn\bar{3}m$	Origin 1	PCC$I3Q666$P4C666$P2D666
224	$Pn\bar{3}m$	Origin 2	PCC$I3Q000$P4C066$P2D660
225	$Fm\bar{3}m$		FCC$I3Q000$P4C000$P2D000
226	$Fm\bar{3}c$		FCC$I3Q000$P4C666$P2D666
227	$Fd\bar{3}m$	Origin 1	FCC$I3Q333$P4C993$P2D939
227	$Fd\bar{3}m$	Origin 2	FCC$I3Q000$P4C693$P2D936
228	$Fd\bar{3}c$	Origin 1	FCC$I3Q999$P4C993$P2D939
228	$Fd\bar{3}c$	Origin 2	FCC$I3Q000$P4C093$P2D930
229	$Im\bar{3}m$		ICC$I3Q000$P4C000$P2D000
230	$Ia\bar{3}d$		ICC$I3Q000$P4C393$P2D933

$$\mathtt{1A}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\quad \mathtt{2A}=\begin{pmatrix}1&0&0\\0&\bar{1}&0\\0&0&\bar{1}\end{pmatrix}\quad \mathtt{2B}=\begin{pmatrix}\bar{1}&0&0\\0&1&0\\0&0&\bar{1}\end{pmatrix}$$

$$\mathtt{2C}=\begin{pmatrix}\bar{1}&0&0\\0&\bar{1}&0\\0&0&1\end{pmatrix}\quad \mathtt{2D}=\begin{pmatrix}0&1&0\\1&0&0\\0&0&\bar{1}\end{pmatrix}\quad \mathtt{2E}=\begin{pmatrix}0&\bar{1}&0\\\bar{1}&0&0\\0&0&\bar{1}\end{pmatrix}$$

$$\mathtt{2F}=\begin{pmatrix}1&\bar{1}&0\\0&\bar{1}&0\\0&0&\bar{1}\end{pmatrix}\quad \mathtt{2G}=\begin{pmatrix}1&0&0\\1&\bar{1}&0\\0&0&\bar{1}\end{pmatrix}\quad \mathtt{3Q}=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$$

$$\mathtt{3C}=\begin{pmatrix}0&\bar{1}&0\\1&\bar{1}&0\\0&0&1\end{pmatrix}\quad \mathtt{4C}=\begin{pmatrix}0&\bar{1}&0\\1&0&0\\0&0&1\end{pmatrix}\quad \mathtt{6C}=\begin{pmatrix}1&\bar{1}&0\\1&0&0\\0&0&1\end{pmatrix},$$

where only matrices of proper rotation are given (and required), since the corresponding matrices of improper rotation are created by the program for appropriate value of the r_i indicator. The first character of a symbol is the order of the axis of rotation and the second character specifies its orientation: in terms of direct-space lattice vectors, we have

$$\mathtt{A}=[100],\ \mathtt{B}=[010],\ \mathtt{C}=[001],\ \mathtt{D}=[110],$$
$$\mathtt{E}=[1\bar{1}0],\ \mathtt{F}=[100],\ \mathtt{G}=[210]\ \text{and}\ \mathtt{Q}=[111]$$

for the standard orientations of the axes of rotation. Note that the axes 2F, 2G, 3C and 6C appear in trigonal and hexagonal space groups.

In the above scheme a space group is determined by one, two or at most three generators [see (A1.4.2.1)]. It should be pointed out that a convenient way of achieving a representation of the space group in any setting and relative to any origin is to start from the standard generators in Table A1.4.2.1 and let the computer program perform the appropriate transformation of the generators only, as in equations (1.4.4.4) and (1.4.4.5). The subsequent expansion of the transformed generators and the formation of the required products [see (A1.4.2.1)

Table A1.4.2.2. *Lattice symbol L*

The lattice symbol L implies Seitz matrices for the lattice translations. For noncentrosymmetric lattices the rotation parts of the Seitz matrices are for *1* (see Table A1.4.2.4). For centrosymmetric lattices the rotation parts are *1* and *−1*. The translation parts in the fourth columns of the Seitz matrices are listed in the last column of the table. The total number of matrices implied by each symbol is given by **nS**.

Noncentrosymmetric		Centrosymmetric		
Symbol	**nS**	Symbol	**nS**	Implied lattice translation(s)
P	1	−P	2	$0,0,0$
A	2	−A	4	$0,0,0 \quad 0,\frac{1}{2},\frac{1}{2}$
B	2	−B	4	$0,0,0 \quad \frac{1}{2},0,\frac{1}{2}$
C	2	−C	4	$0,0,0 \quad \frac{1}{2},\frac{1}{2},0$
I	2	−I	4	$0,0,0 \quad \frac{1}{2},\frac{1}{2},\frac{1}{2}$
R	3	−R	6	$0,0,0 \quad \frac{2}{3},\frac{1}{3},\frac{1}{3} \quad \frac{1}{3},\frac{2}{3},\frac{2}{3}$
H	3	−H	6	$0,0,0 \quad \frac{2}{3},\frac{1}{3},0 \quad \frac{1}{3},\frac{2}{3},0$
F	4	−F	8	$0,0,0 \quad 0,\frac{1}{2},\frac{1}{2} \quad \frac{1}{2},0,\frac{1}{2} \quad \frac{1}{2},\frac{1}{2},0$

and (A1.4.2.2)] leads to the new representation of the space group.

In order to illustrate an explicit space-group symbol consider, for example, the symbol for the space group $Ia\bar{3}d$, as given in Table A1.4.2.1:

```
ICC$I3Q000$P4C393$P2D933.
```

The first three characters tell us that the Bravais lattice of this space group is of type I, that the space group is centrosymmetric and that it belongs to the cubic system. We then see that the generators are (i) an improper threefold axis along [111] (I3Q) with a zero translation part, (ii) a proper fourfold axis along [001] (P4C) with translation part (1/4, 3/4, 1/4) and (iii) a proper twofold axis along [110] (P2D) with translation part (3/4, 1/4, 1/4).

If we make use of the above-outlined interpretation of the explicit symbol (A1.4.2.3), the space-group symmetry transformations in direct space, corresponding to these three generators of the space group $Ia\bar{3}d$, become

$$\left[\begin{pmatrix} 0 & 0 & \bar{1} \\ \bar{1} & 0 & 0 \\ 0 & \bar{1} & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right] = \begin{pmatrix} \bar{z} \\ \bar{x} \\ \bar{y} \end{pmatrix},$$

$$\left[\begin{pmatrix} 0 & \bar{1} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ \frac{3}{4} \\ \frac{1}{4} \end{pmatrix}\right] = \begin{pmatrix} \frac{1}{4}-y \\ \frac{3}{4}+x \\ \frac{1}{4}+z \end{pmatrix},$$

$$\left[\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \bar{1} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}\right] = \begin{pmatrix} \frac{3}{4}+y \\ \frac{1}{4}+x \\ \frac{1}{4}-z \end{pmatrix}.$$

The corresponding symmetry transformations in reciprocal space, in the notation of Section 1.4.4, are

$$\left[(hkl)\begin{pmatrix} 0 & 0 & \bar{1} \\ \bar{1} & 0 & 0 \\ 0 & \bar{1} & 0 \end{pmatrix} : -(hkl)\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right] = [\overline{klh} : 0];$$

similarly, $[k\bar{h}l : -131/4]$ and $[kh\bar{l} : -311/4]$ are obtained from the second and third generator of $Ia\bar{3}d$, respectively.

The first column of Table A1.4.2.1 lists the conventional space-group number. The second column shows the conventional short Hermann–Mauguin or international space-group symbol, and the third column, *Comments*, shows the full international space-group symbol *only* for the different settings of the monoclinic space groups that are given in the main space-group tables of *IT* A (1983). Other comments pertain to the choice of the space-group origin – where there are alternatives – and to axial systems. The fourth column shows the explicit space-group symbols described above for each of the settings considered in *IT* A (1983).

A1.4.2.3. Hall symbols

By S. R. Hall and R. W. Grosse-Kunstleve

The explicit-origin space-group notation proposed by Hall (1981*a*) is based on a subset of the symmetry operations, in the form of Seitz matrices, sufficient to uniquely define a space group. The concise unambiguous nature of this notation makes it well suited to handling symmetry in computing and database applications.

Table A1.4.2.7 lists space-group notation in several formats. The first column of Table A1.4.2.7 lists the space-group numbers with axis codes appended to identify the nonstandard settings. The second column lists the Hermann–Mauguin symbols in computer-entry format with appended codes to identify the origin and cell choice when there are alternatives. The general forms of the Hall notation are listed in the fourth column and the computer-entry representations of these symbols are listed in the third column. The computer-entry format is the general notation expressed as case-insensitive ASCII characters with the overline (bar) symbol replaced by a minus sign.

The Hall notation has the general form:

$$\mathbf{L}[\mathbf{N}_\mathbf{T}^\mathbf{A}]_\mathbf{1} \ldots [\mathbf{N}_\mathbf{T}^\mathbf{A}]_\mathbf{p}\mathbf{V}. \qquad \text{(A1.4.2.4)}$$

Table A1.4.2.3. *Translation symbol T*

The symbol T specifies the translation elements of a Seitz matrix. Alphabetical symbols (given in the first column) specify translations along a fixed direction. Numerical symbols (given in the third column) specify translations as a fraction of the rotation order |N| and in the direction of the implied or explicitly defined axis.

Translation symbol	Translation vector	Subscript symbol	Fractional translation
a	$\frac{1}{2},0,0$	*1* in 3_1	$\frac{1}{3}$
b	$0,\frac{1}{2},0$	*2* in 3_2	$\frac{2}{3}$
c	$0,0,\frac{1}{2}$	*1* in 4_1	$\frac{1}{4}$
n	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$	*3* in 4_3	$\frac{3}{4}$
u	$\frac{1}{4},0,0$	*1* in 6_1	$\frac{1}{6}$
v	$0,\frac{1}{4},0$	*2* in 6_2	$\frac{1}{3}$
w	$0,0,\frac{1}{4}$	*4* in 6_4	$\frac{2}{3}$
d	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$	*5* in 6_5	$\frac{5}{6}$

Table A1.4.2.4. *Rotation matrices for principal axes*

The 3 × 3 matrices for *proper* rotations along the three principal unit-cell directions are given below. The matrices for *improper* rotations (*−1*, *−2*, *−3*, *−4* and *−6*) are identical except that the signs of the elements are reversed.

Axis	Symbol A	Rotation order				
		1	*2*	*3*	*4*	*6*
a	*x*	$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&\bar{1}&0\\0&0&\bar{1}\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&0&\bar{1}\\0&1&\bar{1}\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&0&\bar{1}\\0&1&0\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&1&\bar{1}\\0&1&0\end{pmatrix}$
b	*y*	$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}\bar{1}&0&0\\0&1&0\\0&0&\bar{1}\end{pmatrix}$	$\begin{pmatrix}\bar{1}&0&1\\0&1&0\\\bar{1}&0&0\end{pmatrix}$	$\begin{pmatrix}0&0&1\\0&1&0\\\bar{1}&0&0\end{pmatrix}$	$\begin{pmatrix}0&0&1\\0&1&0\\\bar{1}&0&1\end{pmatrix}$
c	*z*	$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}\bar{1}&0&0\\0&\bar{1}&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}0&\bar{1}&0\\1&\bar{1}&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}0&\bar{1}&0\\1&0&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}1&\bar{1}&0\\1&0&0\\0&0&1\end{pmatrix}$

L is the symbol specifying the lattice translational symmetry (see Table A1.4.2.2). The integral translations are implicitly included in the set of generators. If **L** has a leading minus sign, it also specifies an inversion centre at the origin. $[\mathbf{N}^{\mathbf{A}}_{\mathbf{T}}]_{\mathbf{n}}$ specifies the 4 × 4 Seitz matrix $\mathbf{S}_{\mathbf{n}}$ of a symmetry element in the minimum set which defines the space-group symmetry (see Tables A1.4.2.3 to A1.4.2.6), and **p** is the number of elements in the set. **V** is a change-of-basis operator needed for less common descriptions of the space-group symmetry.

The matrix symbol $\mathbf{N}^{\mathbf{A}}_{\mathbf{T}}$ is composed of three parts: **N** is the symbol denoting the |N|-fold order of the rotation matrix (see Tables A1.4.2.4, A1.4.2.5 and A1.4.2.6), **T** is a subscript symbol denoting the *translation* vector (see Table A1.4.2.3) and **A** is a superscript symbol denoting the *axis* of rotation.

The computer-entry format of the Hall notation contains the rotation-order symbol **N** as positive integers *1*, *2*, *3*, *4*, or *6* for proper rotations and as negative integers *−1*, *−2*, *−3*, *−4* or *−6* for improper rotations. The **T** translation symbols *1*, *2*, *3*, *4*, *5*, *6*, *a*, *b*, *c*, *n*, *u*, *v*, *w*, *d* are described in Table A1.4.2.3. These translations apply additively [*e.g. ad* signifies a $(\frac{3}{4}, \frac{1}{4}, \frac{1}{4})$ translation]. The **A** axis symbols *x*, *y*, *z* denote rotations about the axes **a**, **b** and **c**, respectively (see Table A1.4.2.4). The axis symbols ″ and ′ signal rotations about the body-diagonal vectors **a** + **b** (or alternatively **b** + **c** or **c** + **a**) and **a** − **b** (or alternatively **b** − **c** or **c** − **a**) (see Table A1.4.2.5). The axis symbol * always refers to a threefold rotation along **a** + **b** + **c** (see Table A1.4.2.6).

The change-of-basis operator **V** has the general form (v_x, v_y, v_z). The vectors v_x, v_y and v_z are specified by

$$\begin{aligned} v_x &= r_{1,1}X + r_{1,2}Y + r_{1,3}Z + \mathbf{t}_1 \\ v_y &= r_{2,1}X + r_{2,2}Y + r_{2,3}Z + \mathbf{t}_2\,, \\ v_z &= r_{3,1}X + r_{3,2}Y + r_{3,3}Z + \mathbf{t}_3 \end{aligned}$$

where $r_{i,j}$ and $\mathbf{t}_i$ are fractions or real numbers. Terms in which $r_{i,j}$ or $\mathbf{t}_i$ are zero need not be specified. The 4 × 4 change-of-basis matrix operator **V** is defined as

$$\mathbf{V} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \mathbf{t}_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & \mathbf{t}_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & \mathbf{t}_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The transformed symmetry operations are derived from the specified Seitz matrices $\mathbf{S}_{\mathbf{n}}$ as

$$\mathbf{S}'_{\mathbf{n}} = \mathbf{V} \cdot \mathbf{S}_{\mathbf{n}} \cdot \mathbf{V}^{-1}$$

and from the integral translations **t**(1, 0, 0), **t**(0, 1, 0) and **t**(0, 0, 1) as

$$(\mathbf{t}'_{\mathbf{n}}, \mathbf{1})^T = \mathbf{V} \cdot (\mathbf{t}_{\mathbf{n}}, \mathbf{1})^T.$$

A shorthand form of **V** may be used when the change-of-basis operator only translates the origin of the basis system. In this form v_x, v_y and v_z are specified simply as shifts in twelfths, implying the matrix operator

Table A1.4.2.5. *Rotation matrices for face-diagonal axes*

The symbols for face-diagonal twofold rotations are *2′* and *2″*. The face-diagonal axis direction is determined by the axis of the preceding rotation N^x, N^y or N^z. Note that the single prime ′ is the default and may be omitted.

Preceding rotation	Rotation	Axis	Matrix
N^x	*2′*	**b** − **c**	$\begin{pmatrix}\bar{1}&0&0\\0&0&\bar{1}\\0&\bar{1}&0\end{pmatrix}$
	2″	**b** + **c**	$\begin{pmatrix}\bar{1}&0&0\\0&0&1\\0&1&0\end{pmatrix}$
N^y	*2′*	**a** − **c**	$\begin{pmatrix}0&0&\bar{1}\\0&\bar{1}&0\\\bar{1}&0&0\end{pmatrix}$
	2″	**a** + **c**	$\begin{pmatrix}0&0&1\\0&\bar{1}&0\\1&0&0\end{pmatrix}$
N^z	*2′*	**a** − **b**	$\begin{pmatrix}0&\bar{1}&0\\\bar{1}&0&0\\0&0&\bar{1}\end{pmatrix}$
	2″	**a** + **b**	$\begin{pmatrix}0&1&0\\1&0&0\\0&0&\bar{1}\end{pmatrix}$

Table A1.4.2.6. *Rotation matrix for the body-diagonal axis*

The symbol for the threefold rotation in the **a** + **b** + **c** direction is *3**. Note that for cubic space groups the body-diagonal axis is implied and the asterisk * may be omitted.

Axis	Rotation	Matrix
a + **b** + **c**	*3**	$\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 & v_x/12 \\ 0 & 1 & 0 & v_y/12 \\ 0 & 0 & 1 & v_z/12 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In the shorthand form of **V**, the commas separating the vectors may be omitted.

A1.4.2.3.1. Default axes

For most symbols the rotation axes applicable to each **N** are implied and an explicit axis symbol **A** is not needed. The rules for *default* axis directions are:

(i) the *first* rotation or roto-inversion has an axis direction of **c**;

(ii) the *second* rotation (if |N| is 2) has an axis direction of **a** if preceded by an |N| of 2 or 4, **a**−**b** if preceded by an |N| of 3 or 6;

(iii) the *third* rotation (if |N| is 3) has an axis direction of **a** + **b** + **c**.

A1.4.2.3.2. Example matrices

The following examples show how the notation expands to Seitz matrices.

The notation $\bar{2}_c^x$ represents an improper twofold rotation along **a** and a **c**/2 translation:

$$-2\text{xc} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The notation 3^* represents a threefold rotation along **a** + **b** + **c**:

$$3^* = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The notation 4_{vw} represents a fourfold rotation along **c** (implied) and translation of **b**/4 and **c**/4:

$$4\text{vw} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The notation 6_1 2 (0 0 −1) represents a 6_1 screw along **c**, a twofold rotation along **a** − **b** and an origin shift of −**c**/12. Note that the 6_1 matrix is unchanged by the shifted origin whereas the *2* matrix is changed by −**c**/6.

$$\begin{aligned} &61\,2\,(0\,0\,-1) \\ &\quad = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & \frac{5}{6} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

The change-of-basis vector (0 0 −1) could also be entered as $(x, y, z - 1/12)$.

The *reverse setting of the R-centred lattice* (hexagonal axes) is specified using a change-of-basis transformation applied to the standard *obverse setting* (see Table A1.4.2.2). The obverse Seitz matrices are

$$\text{R}\,3 = \begin{pmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{2}{3} \\ 0 & 0 & 1 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & \frac{2}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The reverse-setting Seitz matrices are

$$\begin{aligned} &\text{R}\,3\,(-\text{x}, -\text{y}, \text{z}) \\ &\quad = \begin{pmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & \frac{2}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

The conventional primitive hexagonal lattice may be transformed to a *C-centred orthohexagonal setting* using the change-of-basis operator

$$\text{P}\,6\,(\text{x} - 1/2\text{y}, 1/2\text{y}, \text{z}) = \begin{pmatrix} \frac{1}{2} & -\frac{3}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In this case the lattice translation for the *C* centring is obtained by transforming the integral translation **t**(0, 1, 0):

$$\begin{aligned} \boldsymbol{V} \cdot \begin{pmatrix} 0 & 1 & 0 & 1 \end{pmatrix}^T &= \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & 0 & 1 \end{pmatrix}^T. \end{aligned}$$

The standard setting of an *I*-centred tetragonal space group may be transformed to a primitive setting using the change-of-basis operator

$$\text{I}\,4\,(\text{y} + \text{z}, \text{x} + \text{z}, \text{x} + \text{y}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that in the primitive setting, the fourfold axis is along **a** + **b**.

Table A1.4.2.7. *Hall symbols*

The first column, n:c, lists the space-group numbers and axis codes† separated by a colon. The second column lists the Hermann–Mauguin symbols in computer-entry format. The third column lists the Hall symbols in computer-entry format and the fourth column lists the Hall symbols as described in Tables A1.4.2.2–A1.4.2.6.

n:c	H–M entry	Hall entry	Hall symbol
1	P 1	p 1	P 1
2	P -1	-p 1	$\overline{\mathrm{P}}$ 1
3:b	P 1 2 1	p 2y	P 2^{y}
3:c	P 1 1 2	p 2	P 2
3:a	P 2 1 1	p 2x	P 2^{x}
4:b	P 1 21 1	p 2yb	P 2_{b}^{y}
4:c	P 1 1 21	p 2c	P 2_{c}
4:a	P 21 1 1	p 2xa	P 2_{a}^{x}
5:b1	C 1 2 1	c 2y	C 2^{y}
5:b2	A 1 2 1	a 2y	A 2^{y}
5:b3	I 1 2 1	i 2y	I 2^{y}
5:c1	A 1 1 2	a 2	A 2
5:c2	B 1 1 2	b 2	B 2
5:c3	I 1 1 2	i 2	I 2
5:a1	B 2 1 1	b 2x	B 2^{x}
5:a2	C 2 1 1	c 2x	C 2^{x}
5:a3	I 2 1 1	i 2x	I 2^{x}
6:b	P 1 m 1	p -2y	P $\overline{2}^{y}$
6:c	P 1 1 m	p -2	P $\overline{2}$
6:a	P m 1 1	p -2x	P $\overline{2}^{x}$
7:b1	P 1 c 1	p -2yc	P $\overline{2}_{c}^{y}$
7:b2	P 1 n 1	p -2yac	P $\overline{2}_{ac}^{y}$
7:b3	P 1 a 1	p -2ya	P $\overline{2}_{a}^{y}$
7:c1	P 1 1 a	p -2a	P $\overline{2}_{a}$
7:c2	P 1 1 n	p -2ab	P $\overline{2}_{ab}$
7:c3	P 1 1 b	p -2b	P $\overline{2}_{b}$
7:a1	P b 1 1	p -2xb	P $\overline{2}_{b}^{x}$
7:a2	P n 1 1	p -2xbc	P $\overline{2}_{bc}^{x}$
7:a3	P c 1 1	p -2xc	P $\overline{2}_{c}^{x}$
8:b1	C 1 m 1	c -2y	C $\overline{2}^{y}$
8:b2	A 1 m 1	a -2y	A $\overline{2}^{y}$
8:b3	I 1 m 1	i -2y	I $\overline{2}^{y}$
8:c1	A 1 1 m	a -2	A $\overline{2}$
8:c2	B 1 1 m	b -2	B $\overline{2}$
8:c3	I 1 1 m	i -2	I $\overline{2}$
8:a1	B m 1 1	b -2x	B $\overline{2}^{x}$
8:a2	C m 1 1	c -2x	C $\overline{2}^{x}$
8:a3	I m 1 1	i -2x	I $\overline{2}^{x}$
9:b1	C 1 c 1	c -2yc	C $\overline{2}_{c}^{y}$
9:b2	A 1 n 1	a -2yab	A $\overline{2}_{ab}^{y}$
9:b3	I 1 a 1	i -2ya	I $\overline{2}_{a}^{y}$
9:-b1	A 1 a 1	a -2ya	A $\overline{2}_{a}^{y}$
9:-b2	C 1 n 1	c -2yac	C $\overline{2}_{ac}^{y}$
9:-b3	I 1 c 1	i -2yc	I $\overline{2}_{c}^{y}$
9:c1	A 1 1 a	a -2a	A $\overline{2}_{a}$
9:c2	B 1 1 n	b -2ab	B $\overline{2}_{ab}$
9:c3	I 1 1 b	i -2b	I $\overline{2}_{b}$
9:-c1	B 1 1 b	b -2b	B $\overline{2}_{b}$
9:-c2	A 1 1 n	a -2ab	A $\overline{2}_{ab}$
9:-c3	I 1 1 a	i -2a	I $\overline{2}_{a}$
9:a1	B b 1 1	b -2xb	B $\overline{2}_{b}^{x}$
9:a2	C n 1 1	c -2xac	C $\overline{2}_{ac}^{x}$
9:a3	I c 1 1	i -2xc	I $\overline{2}_{c}^{x}$
9:-a1	C c 1 1	c -2xc	C $\overline{2}_{c}^{x}$
9:-a2	B n 1 1	b -2xab	B $\overline{2}_{ab}^{x}$
9:-a3	I b 1 1	i -2xb	I $\overline{2}_{b}^{x}$
10:b	P 1 2/m 1	-p 2y	$\overline{\mathrm{P}}$ 2^{y}
10:c	P 1 1 2/m	-p 2	$\overline{\mathrm{P}}$ 2

n:c	H–M entry	Hall entry	Hall symbol
10:a	P 2/m 1 1	-p 2x	$\overline{\mathrm{P}}$ 2^{x}
11:b	P 1 21/m 1	-p 2yb	$\overline{\mathrm{P}}$ 2_{b}^{y}
11:c	P 1 1 21/m	-p 2c	$\overline{\mathrm{P}}$ 2_{c}
11:a	P 21/m 1 1	-p 2xa	$\overline{\mathrm{P}}$ 2_{a}^{x}
12:b1	C 1 2/m 1	-c 2y	$\overline{\mathrm{C}}$ 2^{y}
12:b2	A 1 2/m 1	-a 2y	$\overline{\mathrm{A}}$ 2^{y}
12:b3	I 1 2/m 1	-i 2y	$\overline{\mathrm{I}}$ 2^{y}
12:c1	A 1 1 2/m	-a 2	$\overline{\mathrm{A}}$ 2
12:c2	B 1 1 2/m	-b 2	$\overline{\mathrm{B}}$ 2
12:c3	I 1 1 2/m	-i 2	$\overline{\mathrm{I}}$ 2
12:a1	B 2/m 1 1	-b 2x	$\overline{\mathrm{B}}$ 2^{x}
12:a2	C 2/m 1 1	-c 2x	$\overline{\mathrm{C}}$ 2^{x}
12:a3	I 2/m 1 1	-i 2x	$\overline{\mathrm{I}}$ 2^{x}
13:b1	P 1 2/c 1	-p 2yc	$\overline{\mathrm{P}}$ 2_{c}^{y}
13:b2	P 1 2/n 1	-p 2yac	$\overline{\mathrm{P}}$ 2_{ac}^{y}
13:b3	P 1 2/a 1	-p 2ya	$\overline{\mathrm{P}}$ 2_{a}^{y}
13:c1	P 1 1 2/a	-p 2a	$\overline{\mathrm{P}}$ 2_{a}
13:c2	P 1 1 2/n	-p 2ab	$\overline{\mathrm{P}}$ 2_{ab}
13:c3	P 1 1 2/b	-p 2b	$\overline{\mathrm{P}}$ 2_{b}
13:a1	P 2/b 1 1	-p 2xb	$\overline{\mathrm{P}}$ 2_{b}^{x}
13:a2	P 2/n 1 1	-p 2xbc	$\overline{\mathrm{P}}$ 2_{bc}^{x}
13:a3	P 2/c 1 1	-p 2xc	$\overline{\mathrm{P}}$ 2_{c}^{x}
14:b1	P 1 21/c 1	-p 2ybc	$\overline{\mathrm{P}}$ 2_{bc}^{y}
14:b2	P 1 21/n 1	-p 2yn	$\overline{\mathrm{P}}$ 2_{n}^{y}
14:b3	P 1 21/a 1	-p 2yab	$\overline{\mathrm{P}}$ 2_{ab}^{y}
14:c1	P 1 1 21/a	-p 2ac	$\overline{\mathrm{P}}$ 2_{ac}
14:c2	P 1 1 21/n	-p 2n	$\overline{\mathrm{P}}$ 2_{n}
14:c3	P 1 1 21/b	-p 2bc	$\overline{\mathrm{P}}$ 2_{bc}
14:a1	P 21/b 1 1	-p 2xab	$\overline{\mathrm{P}}$ 2_{ab}^{x}
14:a2	P 21/n 1 1	-p 2xn	$\overline{\mathrm{P}}$ 2_{n}^{x}
14:a3	P 21/c 1 1	-p 2xac	$\overline{\mathrm{P}}$ 2_{ac}^{x}
15:b1	C 1 2/c 1	-c 2yc	$\overline{\mathrm{C}}$ 2_{c}^{y}
15:b2	A 1 2/n 1	-a 2yab	$\overline{\mathrm{A}}$ 2_{ab}^{y}
15:b3	I 1 2/a 1	-i 2ya	$\overline{\mathrm{I}}$ 2_{a}^{y}
15:-b1	A 1 2/a 1	-a 2ya	$\overline{\mathrm{A}}$ 2_{a}^{y}
15:-b2	C 1 2/n 1	-c 2yac	$\overline{\mathrm{C}}$ 2_{ac}^{y}
15:-b3	I 1 2/c 1	-i 2yc	$\overline{\mathrm{I}}$ 2_{c}^{y}
15:c1	A 1 1 2/a	-a 2a	$\overline{\mathrm{A}}$ 2_{a}
15:c2	B 1 1 2/n	-b 2ab	$\overline{\mathrm{B}}$ 2_{ab}
15:c3	I 1 1 2/b	-i 2b	$\overline{\mathrm{I}}$ 2_{b}
15:-c1	B 1 1 2/b	-b 2b	$\overline{\mathrm{B}}$ 2_{b}
15:-c2	A 1 1 2/n	-a 2ab	$\overline{\mathrm{A}}$ 2_{ab}
15:-c3	I 1 1 2/a	-i 2a	$\overline{\mathrm{I}}$ 2_{a}
15:a1	B 2/b 1 1	-b 2xb	$\overline{\mathrm{B}}$ 2_{b}^{x}
15:a2	C 2/n 1 1	-c 2xac	$\overline{\mathrm{C}}$ 2_{ac}^{x}
15:a3	I 2/c 1 1	-i 2xc	$\overline{\mathrm{I}}$ 2_{c}^{x}
15:-a1	C 2/c 1 1	-c 2xc	$\overline{\mathrm{C}}$ 2_{c}^{x}
15:-a2	B 2/n 1 1	-b 2xab	$\overline{\mathrm{B}}$ 2_{ab}^{x}
15:-a3	I 2/b 1 1	-i 2xb	$\overline{\mathrm{I}}$ 2_{b}^{x}
16	P 2 2 2	p 2 2	P 2 2
17	P 2 2 21	p 2c 2	P 2_{c} 2
17:cab	P 21 2 2	p 2a 2a	P 2_{a} 2_{a}
17:bca	P 2 21 2	p 2 2b	P 2 2_{b}
18	P 21 21 2	p 2 2ab	P 2 2_{ab}
18:cab	P 2 21 21	p 2bc 2	P 2_{bc} 2
18:bca	P 21 2 21	p 2ac 2ac	P 2_{ac} 2_{ac}
19	P 21 21 21	p 2ac 2ab	P 2_{ac} 2_{ab}
20	C 2 2 21	c 2c 2	C 2_{c} 2

Table A1.4.2.7 (*cont.*)

n:c	H–M entry	Hall entry	Hall symbol
20:cab	A 21 2 2	a 2a 2a	A 2_a 2_a
20:bca	B 2 21 2	b 2 2b	B 2 2_b
21	C 2 2 2	c 2 2	C 2 2
21:cab	A 2 2 2	a 2 2	A 2 2
21:bca	B 2 2 2	b 2 2	B 2 2
22	F 2 2 2	f 2 2	F 2 2
23	I 2 2 2	i 2 2	I 2 2
24	I 21 21 21	i 2b 2c	I 2_b 2_c
25	P m m 2	p 2 -2	P 2 $\overline{2}$
25:cab	P 2 m m	p -2 2	P $\overline{2}$ 2
25:bca	P m 2 m	p -2 -2	P $\overline{2}$ $\overline{2}$
26	P m c 21	p 2c -2	P 2_c $\overline{2}$
26:ba-c	P c m 21	p 2c -2c	P 2_c $\overline{2}_c$
26:cab	P 21 m a	p -2a 2a	P $\overline{2}_a$ 2_a
26:-cba	P 21 a m	p -2 2a	P $\overline{2}$ 2_a
26:bca	P b 21 m	p -2 -2b	P $\overline{2}$ $\overline{2}_b$
26:a-cb	P m 21 b	p -2b -2	P $\overline{2}_b$ $\overline{2}$
27	P c c 2	p 2 -2c	P 2 $\overline{2}_c$
27:cab	P 2 a a	p -2a 2	P $\overline{2}_a$ 2
27:bca	P b 2 b	p -2b -2b	P $\overline{2}_b$ $\overline{2}_b$
28	P m a 2	p 2 -2a	P 2 $\overline{2}_a$
28:ba-c	P b m 2	p 2 -2b	P 2 $\overline{2}_b$
28:cab	P 2 m b	p -2b 2	P $\overline{2}_b$ 2
28:-cba	P 2 c m	p -2c 2	P $\overline{2}_c$ 2
28:bca	P c 2 m	p -2c -2c	P $\overline{2}_c$ $\overline{2}_c$
28:a-cb	P m 2 a	p -2a -2a	P $\overline{2}_a$ $\overline{2}_a$
29	P c a 21	p 2c -2ac	P 2_c $\overline{2}_{ac}$
29:ba-c	P b c 21	p 2c -2b	P 2_c $\overline{2}_b$
29:cab	P 21 a b	p -2b 2a	P $\overline{2}_b$ 2_a
29:-cba	P 21 c a	p -2ac 2a	P $\overline{2}_{ac}$ 2_a
29:bca	P c 21 b	p -2bc -2c	P $\overline{2}_{bc}$ $\overline{2}_c$
29:a-cb	P b 21 a	p -2a -2ab	P $\overline{2}_a$ $\overline{2}_{ab}$
30	P n c 2	p 2 -2bc	P 2 $\overline{2}_{bc}$
30:ba-c	P c n 2	p 2 -2ac	P 2 $\overline{2}_{ac}$
30:cab	P 2 n a	p -2ac 2	P $\overline{2}_{ac}$ 2
30:-cba	P 2 a n	p -2ab 2	P $\overline{2}_{ab}$ 2
30:bca	P b 2 n	p -2ab -2ab	P $\overline{2}_{ab}$ $\overline{2}_{ab}$
30:a-cb	P n 2 b	p -2bc -2bc	P $\overline{2}_{bc}$ $\overline{2}_{bc}$
31	P m n 21	p 2ac -2	P 2_{ac} $\overline{2}$
31:ba-c	P n m 21	p 2bc -2bc	P 2_{bc} $\overline{2}_{bc}$
31:cab	P 21 m n	p -2ab 2ab	P $\overline{2}_{ab}$ 2_{ab}
31:-cba	P 21 n m	p -2 2ac	P $\overline{2}$ 2_{ac}
31:bca	P n 21 m	p -2 -2bc	P $\overline{2}$ $\overline{2}_{bc}$
31:a-cb	P m 21 n	p -2ab -2	P $\overline{2}_{ab}$ $\overline{2}$
32	P b a 2	p 2 -2ab	P 2 $\overline{2}_{ab}$
32:cab	P 2 c b	p -2bc 2	P $\overline{2}_{bc}$ 2
32:bca	P c 2 a	p -2ac -2ac	P $\overline{2}_{ac}$ $\overline{2}_{ac}$
33	P n a 21	p 2c -2n	P 2_c $\overline{2}_n$
33:ba-c	P b n 21	p 2c -2ab	P 2_c $\overline{2}_{ab}$
33:cab	P 21 n b	p -2bc 2a	P $\overline{2}_{bc}$ 2_a
33:-cba	P 21 c n	p -2n 2a	P $\overline{2}_n$ 2_a
33:bca	P c 21 n	p -2n -2ac	P $\overline{2}_n$ $\overline{2}_{ac}$
33:a-cb	P n 21 a	p -2ac -2n	P $\overline{2}_{ac}$ $\overline{2}_n$
34	P n n 2	p 2 -2n	P 2 $\overline{2}_n$
34:cab	P 2 n n	p -2n 2	P $\overline{2}_n$ 2
34:bca	P n 2 n	p -2n -2n	P $\overline{2}_n$ $\overline{2}_n$
35	C m m 2	c 2 -2	C 2 $\overline{2}$
35:cab	A 2 m m	a -2 2	A $\overline{2}$ 2
35:bca	B m 2 m	b -2 -2	B $\overline{2}$ $\overline{2}$
36	C m c 21	c 2c -2	C 2_c $\overline{2}$

n:c	H–M entry	Hall entry	Hall symbol
36:ba-c	C c m 21	c 2c -2c	C 2_c $\overline{2}_c$
36:cab	A 21 m a	a -2a 2a	A $\overline{2}_a$ 2_a
36:-cba	A 21 a m	a -2 2a	A $\overline{2}$ 2_a
36:bca	B b 21 m	b -2 -2b	B $\overline{2}$ $\overline{2}_b$
36:a-cb	B m 21 b	b -2b -2	B $\overline{2}_b$ $\overline{2}$
37	C c c 2	c 2 -2c	C 2 $\overline{2}_c$
37:cab	A 2 a a	a -2a 2	A $\overline{2}_a$ 2
37:bca	B b 2 b	b -2b -2b	B $\overline{2}_b$ $\overline{2}_b$
38	A m m 2	a 2 -2	A 2 $\overline{2}$
38:ba-c	B m m 2	b 2 -2	B 2 $\overline{2}$
38:cab	B 2 m m	b -2 2	B $\overline{2}$ 2
38:-cba	C 2 m m	c -2 2	C $\overline{2}$ 2
38:bca	C m 2 m	c -2 -2	C $\overline{2}$ $\overline{2}$
38:a-cb	A m 2 m	a -2 -2	A $\overline{2}$ $\overline{2}$
39	A b m 2	a 2 -2b	A 2 $\overline{2}_b$
39:ba-c	B m a 2	b 2 -2a	B 2 $\overline{2}_a$
39:cab	B 2 c m	b -2a 2	B $\overline{2}_a$ 2
39:-cba	C 2 m b	c -2a 2	C $\overline{2}_a$ 2
39:bca	C m 2 a	c -2a -2a	C $\overline{2}_a$ $\overline{2}_a$
39:a-cb	A c 2 m	a -2b -2b	A $\overline{2}_b$ $\overline{2}_b$
40	A m a 2	a 2 -2a	A 2 $\overline{2}_a$
40:ba-c	B b m 2	b 2 -2b	B 2 $\overline{2}_b$
40:cab	B 2 m b	b -2b 2	B $\overline{2}_b$ 2
40:-cba	C 2 c m	c -2c 2	C $\overline{2}_c$ 2
40:bca	C c 2 m	c -2c -2c	C $\overline{2}_c$ $\overline{2}_c$
40:a-cb	A m 2 a	a -2a -2a	A $\overline{2}_a$ $\overline{2}_a$
41	A b a 2	a 2 -2ab	A 2 $\overline{2}_{ab}$
41:ba-c	B b a 2	b 2 -2ab	B 2 $\overline{2}_{ab}$
41:cab	B 2 c b	b -2ab 2	B $\overline{2}_{ab}$ 2
41:-cba	C 2 c b	c -2ac 2	C $\overline{2}_{ac}$ 2
41:bca	C c 2 a	c -2ac -2ac	C $\overline{2}_{ac}$ $\overline{2}_{ac}$
41:a-cb	A c 2 a	a -2ab -2ab	A $\overline{2}_{ab}$ $\overline{2}_{ab}$
42	F m m 2	f 2 -2	F 2 $\overline{2}$
42:cab	F 2 m m	f -2 2	F $\overline{2}$ 2
42:bca	F m 2 m	f -2 -2	F $\overline{2}$ $\overline{2}$
43	F d d 2	f 2 -2d	F 2 $\overline{2}_d$
43:cab	F 2 d d	f -2d 2	F $\overline{2}_d$ 2
43:bca	F d 2 d	f -2d -2d	F $\overline{2}_d$ $\overline{2}_d$
44	I m m 2	i 2 -2	I 2 $\overline{2}$
44:cab	I 2 m m	i -2 2	I $\overline{2}$ 2
44:bca	I m 2 m	i -2 -2	I $\overline{2}$ $\overline{2}$
45	I b a 2	i 2 -2c	I 2 $\overline{2}_c$
45:cab	I 2 c b	i -2a 2	I $\overline{2}_a$ 2
45:bca	I c 2 a	i -2b -2b	I $\overline{2}_b$ $\overline{2}_b$
46	I m a 2	i 2 -2a	I 2 $\overline{2}_a$
46:ba-c	I b m 2	i 2 -2b	I 2 $\overline{2}_b$
46:cab	I 2 m b	i -2b 2	I $\overline{2}_b$ 2
46:-cba	I 2 c m	i -2c 2	I $\overline{2}_c$ 2
46:bca	I c 2 m	i -2c -2c	I $\overline{2}_c$ $\overline{2}_c$
46:a-cb	I m 2 a	i -2a -2a	I $\overline{2}_a$ $\overline{2}_a$
47	P m m m	-p 2 2	$\overline{P}$ 2 2
48:1	P n n n:1	p 2 2 -1n	P 2 2 $\overline{1}_n$
48:2	P n n n:2	-p 2ab 2bc	$\overline{P}$ 2_{ab} 2_{bc}
49	P c c m	-p 2 2c	$\overline{P}$ 2 2_c
49:cab	P m a a	-p 2a 2	$\overline{P}$ 2_a 2
49:bca	P b m b	-p 2b 2b	$\overline{P}$ 2_b 2_b
50:1	P b a n:1	p 2 2 -1ab	P 2 2 $\overline{1}_{ab}$
50:2	P b a n:2	-p 2ab 2b	$\overline{P}$ 2_{ab} 2_b
50:1cab	P n c b:1	p 2 2 -1bc	P 2 2 $\overline{1}_{bc}$
50:2cab	P n c b:2	-p 2b 2bc	$\overline{P}$ 2_b 2_{bc}

Table A1.4.2.7 (*cont.*)

n:c	H–M entry	Hall entry	Hall symbol
50:1bca	P c n a:1	p 2 2 -1ac	P 2 2 $\overline{1}_{ac}$
50:2bca	P c n a:2	-p 2a 2c	$\overline{P}$ 2_a 2_c
51	P m m a	-p 2a 2a	$\overline{P}$ 2_a 2_a
51:ba-c	P m m b	-p 2b 2	$\overline{P}$ 2_b 2
51:cab	P b m m	-p 2 2b	$\overline{P}$ 2 2_b
51:-cba	P c m m	-p 2c 2c	$\overline{P}$ 2_c 2_c
51:bca	P m c m	-p 2c 2	$\overline{P}$ 2_c 2
51:a-cb	P m a m	-p 2 2a	$\overline{P}$ 2 2_a
52	P n n a	-p 2a 2bc	$\overline{P}$ 2_a 2_{bc}
52:ba-c	P n n b	-p 2b 2n	$\overline{P}$ 2_b 2_n
52:cab	P b n n	-p 2n 2b	$\overline{P}$ 2_n 2_b
52:-cba	P c n n	-p 2ab 2c	$\overline{P}$ 2_{ab} 2_c
52:bca	P n c n	-p 2ab 2n	$\overline{P}$ 2_{ab} 2_n
52:a-cb	P n a n	-p 2n 2bc	$\overline{P}$ 2_n 2_{bc}
53	P m n a	-p 2ac 2	$\overline{P}$ 2_{ac} 2
53:ba-c	P n m b	-p 2bc 2bc	$\overline{P}$ 2_{bc} 2_{bc}
53:cab	P b m n	-p 2ab 2ab	$\overline{P}$ 2_{ab} 2_{ab}
53:-cba	P c n m	-p 2 2ac	$\overline{P}$ 2 2_{ac}
53:bca	P n c m	-p 2 2bc	$\overline{P}$ 2 2_{bc}
53:a-cb	P m a n	-p 2ab 2	$\overline{P}$ 2_{ab} 2
54	P c c a	-p 2a 2ac	$\overline{P}$ 2_a 2_{ac}
54:ba-c	P c c b	-p 2b 2c	$\overline{P}$ 2_b 2_c
54:cab	P b a a	-p 2a 2b	$\overline{P}$ 2_a 2_b
54:-cba	P c a a	-p 2ac 2c	$\overline{P}$ 2_{ac} 2_c
54:bca	P b c b	-p 2bc 2b	$\overline{P}$ 2_{bc} 2_b
54:a-cb	P b a b	-p 2b 2ab	$\overline{P}$ 2_b 2_{ab}
55	P b a m	-p 2 2ab	$\overline{P}$ 2 2_{ab}
55:cab	P m c b	-p 2bc 2	$\overline{P}$ 2_{bc} 2
55:bca	P c m a	-p 2ac 2ac	$\overline{P}$ 2_{ac} 2_{ac}
56	P c c n	-p 2ab 2ac	$\overline{P}$ 2_{ab} 2_{ac}
56:cab	P n a a	-p 2ac 2bc	$\overline{P}$ 2_{ac} 2_{bc}
56:bca	P b n b	-p 2bc 2ab	$\overline{P}$ 2_{bc} 2_{ab}
57	P b c m	-p 2c 2b	$\overline{P}$ 2_c 2_b
57:ba-c	P c a m	-p 2c 2ac	$\overline{P}$ 2_c 2_{ac}
57:cab	P m c a	-p 2ac 2a	$\overline{P}$ 2_{ac} 2_a
57:-cba	P m a b	-p 2b 2a	$\overline{P}$ 2_b 2_a
57:bca	P b m a	-p 2a 2ab	$\overline{P}$ 2_a 2_{ab}
57:a-cb	P c m b	-p 2bc 2c	$\overline{P}$ 2_{bc} 2_c
58	P n n m	-p 2 2n	$\overline{P}$ 2 2_n
58:cab	P m n n	-p 2n 2	$\overline{P}$ 2_n 2
58:bca	P n m n	-p 2n 2n	$\overline{P}$ 2_n 2_n
59:1	P m m n:1	p 2 2ab -1ab	P 2 2_{ab} $\overline{1}_{ab}$
59:2	P m m n:2	-p 2ab 2a	$\overline{P}$ 2_{ab} 2_a
59:1cab	P n m m:1	p 2bc 2 -1bc	P 2_{bc} 2 $\overline{1}_{bc}$
59:2cab	P n m m:2	-p 2c 2bc	$\overline{P}$ 2_c 2_{bc}
59:1bca	P m n m:1	p 2ac 2ac -1ac	P 2_{ac} 2_{ac} $\overline{1}_{ac}$
59:2bca	P m n m:2	-p 2c 2a	$\overline{P}$ 2_c 2_a
60	P b c n	-p 2n 2ab	$\overline{P}$ 2_n 2_{ab}
60:ba-c	P c a n	-p 2n 2c	$\overline{P}$ 2_n 2_c
60:cab	P n c a	-p 2a 2n	$\overline{P}$ 2_a 2_n
60:-cba	P n a b	-p 2bc 2n	$\overline{P}$ 2_{bc} 2_n
60:bca	P b n a	-p 2ac 2b	$\overline{P}$ 2_{ac} 2_b
60:a-cb	P c n b	-p 2b 2ac	$\overline{P}$ 2_b 2_{ac}
61	P b c a	-p 2ac 2ab	$\overline{P}$ 2_{ac} 2_{ab}
61:ba-c	P c a b	-p 2bc 2ac	$\overline{P}$ 2_{bc} 2_{ac}
62	P n m a	-p 2ac 2n	$\overline{P}$ 2_{ac} 2_n
62:ba-c	P m n b	-p 2bc 2a	$\overline{P}$ 2_{bc} 2_a
62:cab	P b n m	-p 2c 2ab	$\overline{P}$ 2_c 2_{ab}
62:-cba	P c m n	-p 2n 2ac	$\overline{P}$ 2_n 2_{ac}
62:bca	P m c n	-p 2n 2a	$\overline{P}$ 2_n 2_a

n:c	H–M entry	Hall entry	Hall symbol
62:a-cb	P n a m	-p 2c 2n	$\overline{P}$ 2_c 2_n
63	C m c m	-c 2c 2	$\overline{C}$ 2_c 2
63:ba-c	C c m m	-c 2c 2c	$\overline{C}$ 2_c 2_c
63:cab	A m m a	-a 2a 2a	$\overline{A}$ 2_a 2_a
63:-cba	A m a m	-a 2 2a	$\overline{A}$ 2 2_a
63:bca	B b m m	-b 2 2b	$\overline{B}$ 2 2_b
63:a-cb	B m m b	-b 2b 2	$\overline{B}$ 2_b 2
64	C m c a	-c 2ac 2	$\overline{C}$ 2_{ac} 2
64:ba-c	C c m b	-c 2ac 2ac	$\overline{C}$ 2_{ac} 2_{ac}
64:cab	A b m a	-a 2ab 2ab	$\overline{A}$ 2_{ab} 2_{ab}
64:-cba	A c a m	-a 2 2ab	$\overline{A}$ 2 2_{ab}
64:bca	B b c m	-b 2 2ab	$\overline{B}$ 2 2_{ab}
64:a-cb	B m a b	-b 2ab 2	$\overline{B}$ 2_{ab} 2
65	C m m m	-c 2 2	$\overline{C}$ 2 2
65:cab	A m m m	-a 2 2	$\overline{A}$ 2 2
65:bca	B m m m	-b 2 2	$\overline{B}$ 2 2
66	C c c m	-c 2 2c	$\overline{C}$ 2 2_c
66:cab	A m a a	-a 2a 2	$\overline{A}$ 2_a 2
66:bca	B b m b	-b 2b 2b	$\overline{B}$ 2_b 2_b
67	C m m a	-c 2a 2	$\overline{C}$ 2_a 2
67:ba-c	C m m b	-c 2a 2a	$\overline{C}$ 2_a 2_a
67:cab	A b m m	-a 2b 2b	$\overline{A}$ 2_b 2_b
67:-cba	A c m m	-a 2 2b	$\overline{A}$ 2 2_b
67:bca	B m c m	-b 2 2a	$\overline{B}$ 2 2_a
67:a-cb	B m a m	-b 2a 2	$\overline{B}$ 2_a 2
68:1	C c c a:1	c 2 2 -1ac	C 2 2 $\overline{1}_{ac}$
68:2	C c c a:2	-c 2a 2ac	$\overline{C}$ 2_a 2_{ac}
68:1ba-c	C c c b:1	c 2 2 -1ac	C 2 2 $\overline{1}_{ac}$
68:2ba-c	C c c b:2	-c 2a 2c	$\overline{C}$ 2_a 2_c
68:1cab	A b a a:1	a 2 2 -1ab	A 2 2 $\overline{1}_{ab}$
68:2cab	A b a a:2	-a 2a 2b	$\overline{A}$ 2_a 2_b
68:1-cba	A c a a:1	a 2 2 -1ab	A 2 2 $\overline{1}_{ab}$
68:2-cba	A c a a:2	-a 2ab 2b	$\overline{A}$ 2_{ab} 2_b
68:1bca	B b c b:1	b 2 2 -1ab	B 2 2 $\overline{1}_{ab}$
68:2bca	B b c b:2	-b 2ab 2b	$\overline{B}$ 2_{ab} 2_b
68:1a-cb	B b a b:1	b 2 2 -1ab	B 2 2 $\overline{1}_{ab}$
68:2a-cb	B b a b:2	-b 2b 2ab	$\overline{B}$ 2_b 2_{ab}
69	F m m m	-f 2 2	$\overline{F}$ 2 2
70:1	F d d d:1	f 2 2 -1d	F 2 2 $\overline{1}_d$
70:2	F d d d:2	-f 2uv 2vw	$\overline{F}$ 2_{uv} 2_{vw}
71	I m m m	-i 2 2	$\overline{I}$ 2 2
72	I b a m	-i 2 2c	$\overline{I}$ 2 2_c
72:cab	I m c b	-i 2a 2	$\overline{I}$ 2_a 2
72:bca	I c m a	-i 2b 2b	$\overline{I}$ 2_b 2_b
73	I b c a	-i 2b 2c	$\overline{I}$ 2_b 2_c
73:ba-c	I c a b	-i 2a 2b	$\overline{I}$ 2_a 2_b
74	I m m a	-i 2b 2	$\overline{I}$ 2_b 2
74:ba-c	I m m b	-i 2a 2a	$\overline{I}$ 2_a 2_a
74:cab	I b m m	-i 2c 2c	$\overline{I}$ 2_c 2_c
74:-cba	I c m m	-i 2 2b	$\overline{I}$ 2 2_b
74:bca	I m c m	-i 2 2a	$\overline{I}$ 2 2_a
74:a-cb	I m a m	-i 2c 2	$\overline{I}$ 2_c 2
75	P 4	p 4	P 4
76	P 41	p 4w	P 4_w
77	P 42	p 4c	P 4_c
78	P 43	p 4cw	P 4_{cw}
79	I 4	i 4	I 4
80	I 41	i 4bw	I 4_{bw}
81	P -4	p -4	P $\overline{4}$
82	I -4	i -4	I $\overline{4}$

Table A1.4.2.7 (*cont.*)

n:c	H–M entry	Hall entry	Hall symbol
83	P 4/m	-p 4	$\overline{\mathrm{P}}$ 4
84	P 42/m	-p 4c	$\overline{\mathrm{P}}$ 4_{c}
85:1	P 4/n:1	p 4ab -1ab	P 4_{ab} $\overline{1}_{ab}$
85:2	P 4/n:2	-p 4a	$\overline{\mathrm{P}}$ 4_{a}
86:1	P 42/n:1	p 4n -1n	P 4_{n} $\overline{1}_{n}$
86:2	P 42/n:2	-p 4bc	$\overline{\mathrm{P}}$ 4_{bc}
87	I 4/m	-i 4	$\overline{\mathrm{I}}$ 4
88:1	I 41/a:1	i 4bw -1bw	I 4_{bw} $\overline{1}_{bw}$
88:2	I 41/a:2	-i 4ad	$\overline{\mathrm{I}}$ 4_{ad}
89	P 4 2 2	p 4 2	P 4 2
90	P 4 21 2	p 4ab 2ab	P 4_{ab} 2_{ab}
91	P 41 2 2	p 4w 2c	P 4_{w} 2_{c}
92	P 41 21 2	p 4abw 2nw	P$4_{abw}2_{nw}$
93	P 42 2 2	p 4c 2	P 4_{c} 2
94	P 42 21 2	p 4n 2n	P 4_{n} 2_{n}
95	P 43 2 2	p 4cw 2c	P 4_{cw} 2_{c}
96	P 43 21 2	p 4nw 2abw	P 4_{nw} 2_{abw}
97	I 4 2 2	i 4 2	I 4 2
98	I 41 2 2	i 4bw 2bw	I 4_{bw} 2_{bw}
99	P 4 m m	p 4 -2	P 4 $\overline{2}$
100	P 4 b m	p 4 -2ab	P 4 $\overline{2}_{ab}$
101	P 42 c m	p 4c -2c	P 4_{c} $\overline{2}_{c}$
102	P 42 n m	p 4n -2n	P 4_{n} $\overline{2}_{n}$
103	P 4 c c	p 4 -2c	P 4 $\overline{2}_{c}$
104	P 4 n c	p 4 -2n	P 4 $\overline{2}_{n}$
105	P 42 m c	p 4c -2	P 4_{c} $\overline{2}$
106	P 42 b c	p 4c -2ab	P 4_{c} $\overline{2}_{ab}$
107	I 4 m m	i 4 -2	I 4 $\overline{2}$
108	I 4 c m	i 4 -2c	I 4 $\overline{2}_{c}$
109	I 41 m d	i 4bw -2	I 4_{bw} $\overline{2}$
110	I 41 c d	i 4bw -2c	I 4_{bw} $\overline{2}_{c}$
111	P -4 2 m	p -4 2	P $\overline{4}$ 2
112	P -4 2 c	p -4 2c	P $\overline{4}$ 2_{c}
113	P -4 21 m	p -4 2ab	P $\overline{4}$ 2_{ab}
114	P -4 21 c	p -4 2n	P $\overline{4}$ 2_{n}
115	P -4 m 2	p -4 -2	P $\overline{4}$ $\overline{2}$
116	P -4 c 2	p -4 -2c	P $\overline{4}$ $\overline{2}_{c}$
117	P -4 b 2	p -4 -2ab	P $\overline{4}$ $\overline{2}_{ab}$
118	P -4 n 2	p -4 -2n	P $\overline{4}$ $\overline{2}_{n}$
119	I -4 m 2	i -4 -2	I $\overline{4}$ $\overline{2}$
120	I -4 c 2	i -4 -2c	I $\overline{4}$ $\overline{2}_{c}$
121	I -4 2 m	i -4 2	I $\overline{4}$ 2
122	I -4 2 d	i -4 2bw	I $\overline{4}$ 2_{bw}
123	P 4/m m m	-p 4 2	$\overline{\mathrm{P}}$ 4 2
124	P 4/m c c	-p 4 2c	$\overline{\mathrm{P}}$ 4 2_{c}
125:1	P 4/n b m:1	p 4 2 -1ab	P 4 2 $\overline{1}_{ab}$
125:2	P 4/n b m:2	-p 4a 2b	$\overline{\mathrm{P}}$ 4_{a} 2_{b}
126:1	P 4/n n c:1	p 4 2 -1n	P 4 2 $\overline{1}_{n}$
126:2	P 4/n n c:2	-p 4a 2bc	$\overline{\mathrm{P}}$ 4_{a} 2_{bc}
127	P 4/m b m	-p 4 2ab	$\overline{\mathrm{P}}$ 4 2_{ab}
128	P 4/m n c	-p 4 2n	$\overline{\mathrm{P}}$ 4 2_{n}
129:1	P 4/n m m:1	p 4ab 2ab -1ab	P 4_{ab} 2_{ab} $\overline{1}_{ab}$
129:2	P 4/n m m:2	-p 4a 2a	$\overline{\mathrm{P}}$ 4_{a} 2_{a}
130:1	P 4/n c c:1	p 4ab 2n -1ab	P 4_{ab} 2_{n} $\overline{1}_{ab}$
130:2	P 4/n c c:2	-p 4a 2ac	$\overline{\mathrm{P}}$ 4_{a} 2_{ac}
131	P 42/m m c	-p 4c 2	$\overline{\mathrm{P}}$ 4_{c} 2
132	P 42/m c m	-p 4c 2c	$\overline{\mathrm{P}}$ 4_{c} 2_{c}
133:1	P 42/n b c:1	p 4n 2c -1n	P 4_{n} 2_{c} $\overline{1}_{n}$
133:2	P 42/n b c:2	-p 4ac 2b	$\overline{\mathrm{P}}$ 4_{ac} 2_{b}
134:1	P 42/n n m:1	p 4n 2 -1n	P 4_{n} 2 $\overline{1}_{n}$
134:2	P 42/n n m:2	-p 4ac 2bc	$\overline{\mathrm{P}}$ 4_{ac} 2_{bc}
135	P 42/m b c	-p 4c 2ab	$\overline{\mathrm{P}}$ 4_{c} 2_{ab}
136	P 42/m n m	-p 4n 2n	$\overline{\mathrm{P}}$ 4_{n} 2_{n}
137:1	P 42/n m c:1	p 4n 2n -1n	P 4_{n} 2_{n} $\overline{1}_{n}$
137:2	P 42/n m c:2	-p 4ac 2a	$\overline{\mathrm{P}}$ 4_{ac} 2_{a}
138:1	P 42/n c m:1	p 4n 2ab -1n	P 4_{n} 2_{ab} $\overline{1}_{n}$
138:2	P 42/n c m:2	-p 4ac 2ac	$\overline{\mathrm{P}}$ 4_{ac} 2_{ac}
139	I 4/m m m	-i 4 2	$\overline{\mathrm{I}}$ 4 2
140	I 4/m c m	-i 4 2c	$\overline{\mathrm{I}}$ 4 2_{c}
141:1	I 41/a m d:1	i 4bw 2bw -1bw	I 4_{bw} 2_{bw} $\overline{1}_{bw}$
141:2	I 41/a m d:2	-i 4bd 2	$\overline{\mathrm{I}}$ 4_{bd} 2
142:1	I 41/a c d:1	i 4bw 2aw -1bw	I 4_{bw} 2_{aw} $\overline{1}_{bw}$
142:2	I 41/a c d:2	-i 4bd 2c	$\overline{\mathrm{I}}$ 4_{bd} 2_{c}
143	P 3	p 3	P 3
144	P 31	p 31	P 3_{1}
145	P 32	p 32	P 3_{2}
146:h	R 3:h	r 3	R 3
146:r	R 3:r	p 3*	P 3*
147	P -3	-p 3	$\overline{\mathrm{P}}$ 3
148:h	R -3:h	-r 3	$\overline{\mathrm{R}}$ 3
148:r	R -3:r	-p 3*	$\overline{\mathrm{P}}$ 3*
149	P 3 1 2	p 3 2	P 3 2
150	P 3 2 1	p 3 2"	P 3 2"
151	P 31 1 2	p 31 2 (0 0 4)	P 3_{1} 2 (0 0 4)
152	P 31 2 1	p 31 2"	P 3_{1} 2"
153	P 32 1 2	p 32 2 (0 0 2)	P 3_{2} 2 (0 0 2)
154	P 32 2 1	p 32 2"	P 3_{2} 2"
155:h	R 3 2:h	r 3 2"	R 3 2"
155:r	R 3 2:r	p 3* 2	P 3* 2
156	P 3 m 1	p 3 -2"	P 3 $\overline{2}$"
157	P 3 1 m	p 3 -2	P 3 $\overline{2}$
158	P 3 c 1	p 3 -2"c	P 3 $\overline{2}$"$_{c}$
159	P 3 1 c	p 3 -2c	P 3 $\overline{2}_{c}$
160:h	R 3 m:h	r 3 -2"	R 3 $\overline{2}$"
160:r	R 3 m:r	p 3* -2	P 3* $\overline{2}$
161:h	R 3 c:h	r 3 -2"c	R 3 $\overline{2}$"$_{c}$
161:r	R 3 c:r	p 3* -2n	P 3* $\overline{2}_{n}$
162	P -3 1 m	-p 3 2	$\overline{\mathrm{P}}$ 3 2
163	P -3 1 c	-p 3 2c	$\overline{\mathrm{P}}$ 3 2_{c}
164	P -3 m 1	-p 3 2"	$\overline{\mathrm{P}}$ 3 2"
165	P -3 c 1	-p 3 2"c	$\overline{\mathrm{P}}$ 3 2"$_{c}$
166:h	R -3 m:h	-r 3 2"	$\overline{\mathrm{R}}$ 3 2"
166:r	R -3 m:r	-p 3* 2	$\overline{\mathrm{P}}$ 3* 2
167:h	R -3 c:h	-r 3 2"c	$\overline{\mathrm{R}}$ 3 2"$_{c}$
167:r	R -3 c:r	-p 3* 2n	$\overline{\mathrm{P}}$ 3* 2_{n}
168	P 6	p 6	P 6
169	P 61	p 61	P 6_{1}
170	P 65	p 65	P 6_{5}
171	P 62	p 62	P 6_{2}
172	P 64	p 64	P 6_{4}
173	P 63	p 6c	P 6_{c}
174	P -6	p -6	P $\overline{6}$
175	P 6/m	-p 6	$\overline{\mathrm{P}}$ 6
176	P 63/m	-p 6c	$\overline{\mathrm{P}}$ 6_{c}
177	P 6 2 2	p 6 2	P 6 2
178	P 61 2 2	p 61 2 (0 0 5)	P 6_{1} 2 (0 0 5)
179	P 65 2 2	p 65 2 (0 0 1)	P 6_{5} 2 (0 0 1)
180	P 62 2 2	p 62 2 (0 0 4)	P 6_{2} 2 (0 0 4)
181	P 64 2 2	p 64 2 (0 0 2)	P 6_{4} 2 (0 0 2)
182	P 63 2 2	p 6c 2c	P 6_{c} 2_{c}

Table A1.4.2.7 (*cont.*)

n:c	H–M entry	Hall entry	Hall symbol
183	P 6 m m	p 6 -2	P 6 $\overline{2}$
184	P 6 c c	p 6 -2c	P 6 $\overline{2}_c$
185	P 63 c m	p 6c -2	P 6_c $\overline{2}$
186	P 63 m c	p 6c -2c	P 6_c $\overline{2}_c$
187	P -6 m 2	p -6 2	P $\overline{6}$ 2
188	P -6 c 2	p -6c 2	P $\overline{6}_c$ 2
189	P -6 2 m	p -6 -2	P $\overline{6}$ $\overline{2}$
190	P -6 2 c	p -6c -2c	P $\overline{6}_c$ $\overline{2}_c$
191	P 6/m m m	-p 6 2	$\overline{P}$ 6 2
192	P 6/m c c	-p 6 2c	$\overline{P}$ 6 2_c
193	P 63/m c m	-p 6c 2	$\overline{P}$ 6_c 2
194	P 63/m m c	-p 6c 2c	$\overline{P}$ 6_c 2_c
195	P 2 3	p 2 2 3	P 2 2 3
196	F 2 3	f 2 2 3	F 2 2 3
197	I 2 3	i 2 2 3	I 2 2 3
198	P 21 3	p 2ac 2ab 3	P 2_{ac} 2_{ab} 3
199	I 21 3	i 2b 2c 3	I 2_b 2_c 3
200	P m -3	-p 2 2 3	$\overline{P}$ 2 2 3
201:1	P n -3:1	p 2 2 3 -1n	P 2 2 3 $\overline{1}_n$
201:2	P n -3:2	-p 2ab 2bc 3	$\overline{P}$ 2_{ab} 2_{bc} 3
202	F m -3	-f 2 2 3	$\overline{F}$ 2 2 3
203:1	F d -3:1	f 2 2 3 -1d	F 2 2 3 $\overline{1}_d$
203:2	F d -3:2	-f 2uv 2vw 3	$\overline{F}$ 2_{uv} 2_{vw} 3
204	I m -3	-i 2 2 3	$\overline{I}$ 2 2 3
205	P a -3	-p 2ac 2ab 3	$\overline{P}$ 2_{ac} 2_{ab} 3
206	I a -3	-i 2b 2c 3	$\overline{I}$ 2_b 2_c 3
207	P 4 3 2	p 4 2 3	P 4 2 3
208	P 42 3 2	p 4n 2 3	P 4_n 2 3
209	F 4 3 2	f 4 2 3	F 4 2 3
210	F 41 3 2	f 4d 2 3	F 4_d 2 3
211	I 4 3 2	i 4 2 3	I 4 2 3
212	P 43 3 2	p 4acd 2ab 3	P 4_{acd} 2_{ab} 3
213	P 41 3 2	p 4bd 2ab 3	P 4_{bd} 2_{ab} 3
214	I 41 3 2	i 4bd 2c 3	I 4_{bd} 2_c 3
215	P -4 3 m	p -4 2 3	P $\overline{4}$ 2 3
216	F -4 3 m	f -4 2 3	F $\overline{4}$ 2 3
217	I -4 3 m	i -4 2 3	I $\overline{4}$ 2 3
218	P -4 3 n	p -4n 2 3	P $\overline{4}_n$ 2 3
219	F -4 3 c	f -4a 2 3	F $\overline{4}_a$ 2 3
220	I -4 3 d	i -4bd 2c 3	I $\overline{4}_{bd}$ 2_c 3
221	P m -3 m	-p 4 2 3	$\overline{P}$ 4 2 3
222:1	P n -3 n:1	p 4 2 3 -1n	P 4 2 3 $\overline{1}_n$
222:2	P n -3 n:2	-p 4a 2bc 3	$\overline{P}$ 4_a 2_{bc} 3
223	P m -3 n	-p 4n 2 3	$\overline{P}$ 4_n 2 3
224:1	P n -3 m:1	p 4n 2 3 -1n	P 4_n 2 3 $\overline{1}_n$
224:2	P n -3 m:2	-p 4bc 2bc 3	$\overline{P}$ 4_{bc} 2_{bc} 3
225	F m -3 m	-f 4 2 3	$\overline{F}$ 4 2 3
226	F m -3 c	-f 4a 2 3	$\overline{F}$ 4_a 2 3
227:1	F d -3 m:1	f 4d 2 3 -1d	F 4_d 2 3 $\overline{1}_d$
227:2	F d -3 m:2	-f 4vw 2vw 3	$\overline{F}$ 4_{vw} 2_{vw} 3
228:1	F d -3 c:1	f 4d 2 3 -1ad	F 4_d 2 3 $\overline{1}_{ad}$
228:2	F d -3 c:2	-f 4ud 2vw 3	$\overline{F}$ 4_{ud} 2_{vw} 3
229	I m -3 m	-i 4 2 3	$\overline{I}$ 4 2 3
230	I a -3 d	-i 4bd 2c 3	$\overline{I}$ 4_{bd} 2_c 3

The codes appended to the space-group numbers listed in the first column identify the relationship between the symmetry elements and the crystal cell. Where no code is given the first choice listed below applies. *Monoclinic*. Code = *<unique axis><cell choice>*: unique axis choices [*cf. IT* A (2005) Table 4.3.2.1] b, -b, c, -c, a, -a; cell choices [*cf. IT* A (2005) Table 4.3.2.1] 1, 2, 3. *Orthorhombic*. Code = *<origin choice><setting>*: origin choices 1, 2; setting choices [*cf. IT* A (2005) Table 4.3.2.1] abc, ba-c, cab, -cba, bca, a-cb. *Tetragonal, cubic*. Code = *<origin choice>*: origin choices 1, 2. *Trigonal*. Code = *<cell choice>*: cell choices h (hexagonal), r (rhombohedral).

APPENDIX A1.4.3
Structure-factor tables

Table A1.4.3.1. *Plane groups*

The symbols appearing in this table are explained in Section 1.4.3 and in Tables A1.4.3.3 (monoclinic), A1.4.3.5 (tetragonal) and A1.4.3.6 (trigonal and hexagonal).

System	No.	Symbol	Parity	A	B
Oblique	1	$p1$		c(hk)	s(hk)
	2	$p2$		2c(hk)	0
Rectangular	3	pm		2c(hx)c(ky)	2c(hx)s(ky)
	4	pg	$k = 2n$	2c(hx)c(ky)	2c(hx)s(ky)
			$k = 2n + 1$	−2s(hx)s(ky)	2s(hx)c(ky)
	5	cm		4c(hx)c(ky)	4c(hx)s(ky)
	6	$p2mm$		4c(hx)c(ky)	0
	7	$p2mg$	$h = 2n$	4c(hx)c(ky)	0
			$h = 2n + 1$	−4s(hx)s(ky)	0
	8	$p2gg$	$h + k = 2n$	4c(hx)c(ky)	0
			$h + k = 2n + 1$	−4s(hx)s(ky)	0
	9	$c2mm$		8c(hx)c(ky)	0
Square	10	$p4$		2[P(cc) − M(ss)]	0
	11	$p4mm$		4P(cc)	0
	12	$p4gm$	$h + k = 2n$	4P(cc)	0
			$h + k = 2n + 1$	−4M(ss)	0
Hexagonal	13	$p3$		C(hki)	S(hki)
	14	$p3m1$		PH(cc)	MH(ss)
	15	$p31m$		PH(cc)	PH(ss)
	16	$p6$		2C(hki)	0
	17	$p6mm$		2PH(cc)	0

Table A1.4.3.2. *Triclinic space groups*

For the definition of the triple products ccc, csc *etc*., see Table A1.4.3.4.

$P1$ [No. 1]

hkl	A	B
All	$\cos 2\pi(hx + ky + lz)$ = ccc − css − scs − ssc	$\sin 2\pi(hx + ky + lz)$ = scc + csc + ccs − sss

$P\bar{1}$ [No. 2]

hkl	A	B
All	2(ccc − css − scs − ssc)	0

Table A1.4.3.3. *Monoclinic space groups*

Each expression for A or B in the monoclinic system and for the space-group settings chosen in *IT* A is represented in terms of one of the following symbols:

$$\begin{aligned}
&\mathrm{c}(hl)\mathrm{c}(ky) = \cos[2\pi(hx+lz)]\cos(2\pi ky), \quad && \mathrm{c}(hk)\mathrm{c}(lz) = \cos[2\pi(hx+ky)]\cos(2\pi lz),\\
&\mathrm{c}(hl)\mathrm{s}(ky) = \cos[2\pi(hx+lz)]\sin(2\pi ky), \quad && \mathrm{c}(hk)\mathrm{s}(lz) = \cos[2\pi(hx+ky)]\sin(2\pi lz),\\
&\mathrm{s}(hl)\mathrm{c}(ky) = \sin[2\pi(hx+lz)]\cos(2\pi ky), \quad && \mathrm{s}(hk)\mathrm{c}(lz) = \sin[2\pi(hx+ky)]\cos(2\pi lz),\\
&\mathrm{s}(hl)\mathrm{s}(ky) = \sin[2\pi(hx+lz)]\sin(2\pi ky), \quad && \mathrm{s}(hk)\mathrm{s}(lz) = \sin[2\pi(hx+ky)]\sin(2\pi lz), \qquad \text{(A1.4.3.1)}
\end{aligned}$$

where the left-hand column of expressions corresponds to space-group representations in the second setting, with b taken as the unique axis, and the right-hand column corresponds to representations in the first setting, with c taken as the unique axis.

The lattice types in this table are P, A, B, C and I, and are all explicit in the full space-group symbol only (see below). Note that s(hl), s(hk), s(ky) and s(lz) are zero for $h = l = 0$, $h = k = 0$, $k = 0$ and $l = 0$, respectively.

No.	Group symbol		Parity	A	B	Unique axis
	Short	Full				
3	$P2$	$P121$		2c(hl)c(ky)	2c(hl)s(ky)	b
3	$P2$	$P112$		2c(hk)c(lz)	2c(hk)s(lz)	c
4	$P2_1$	$P12_11$	$k = 2n$	2c(hl)c(ky)	2c(hl)s(ky)	b
			$k = 2n+1$	−2s(hl)s(ky)	2s(hl)c(ky)	
4	$P2_1$	$P112_1$	$l = 2n$	2c(hk)c(lz)	2c(hk)s(lz)	c
			$l = 2n+1$	−2s(hk)s(lz)	2s(hk)c(lz)	
5	$C2$	$C121$		4c(hl)c(ky)	4c(hl)s(ky)	b
5	$C2$	$A121$		4c(hl)c(ky)	4c(hl)s(ky)	b
5	$C2$	$I121$		4c(hl)c(ky)	4c(hl)s(ky)	b
5	$C2$	$A112$		4c(hk)c(lz)	4c(hk)s(lz)	c
5	$C2$	$B112$		4c(hk)c(lz)	4c(hk)s(lz)	c
5	$C2$	$I112$		4c(hk)c(lz)	4c(hk)s(lz)	c
6	Pm	$P1m1$		2c(hl)c(ky)	2s(hl)c(ky)	b
6	Pm	$P11m$		2c(hk)c(lz)	2s(hk)c(lz)	c
7	Pc	$P1c1$	$l = 2n$	2c(hl)c(ky)	2s(hl)c(ky)	b
			$l = 2n+1$	−2s(hl)s(ky)	2c(hl)s(ky)	
7	Pc	$P1n1$	$h+l = 2n$	2c(hl)c(ky)	2s(hl)c(ky)	b
			$h+l = 2n+1$	−2s(hl)s(ky)	2c(hl)s(ky)	
7	Pc	$P1a1$	$h = 2n$	2c(hl)c(ky)	2s(hl)c(ky)	b
			$h = 2n+1$	−2s(hl)s(ky)	2c(hl)s(ky)	
7	Pc	$P11a$	$h = 2n$	2c(hk)c(lz)	2s(hk)c(lz)	c
			$h = 2n+1$	−2s(hk)s(lz)	2c(hk)s(lz)	
7	Pc	$P11n$	$h+k = 2n$	2c(hk)c(lz)	2s(hk)c(lz)	c
			$h+k = 2n+1$	−2s(hk)s(lz)	2c(hk)s(lz)	
7	Pc	$P11b$	$k = 2n$	2c(hk)c(lz)	2s(hk)c(lz)	c
			$k = 2n+1$	−2s(hk)s(lz)	2c(hk)s(lz)	
8	Cm	$C1m1$		4c(hl)c(ky)	4s(hl)c(ky)	b
8	Cm	$A1m1$		4c(hl)c(ky)	4s(hl)c(ky)	b
8	Cm	$I1m1$		4c(hl)c(ky)	4s(hl)c(ky)	b
8	Cm	$A11m$		4c(hk)c(lz)	4s(hk)c(lz)	c
8	Cm	$B11m$		4c(hk)c(lz)	4s(hk)c(lz)	c
8	Cm	$I11m$		4c(hk)c(lz)	4s(hk)c(lz)	c
9	Cc	$C1c1$	$l = 2n$	4c(hl)c(ky)	4s(hl)c(ky)	b
			$l = 2n+1$	−4s(hl)s(ky)	4c(hl)s(ky)	
9	Cc	$A1n1$	$h+l = 2n$	4c(hl)c(ky)	4s(hl)c(ky)	b
			$h+l = 2n+1$	−4s(hl)s(ky)	4c(hl)s(ky)	
9	Cc	$I1a1$	$h = 2n$	4c(hl)c(ky)	4s(hl)c(ky)	b
			$h = 2n+1$	−4s(hl)s(ky)	4c(hl)s(ky)	
9	Cc	$A11a$	$h = 2n$	4c(hk)c(lz)	4s(hk)c(lz)	c
			$h = 2n+1$	−4s(hk)s(lz)	4c(hk)s(lz)	
9	Cc	$B11n$	$h+k = 2n$	4c(hk)c(lz)	4s(hk)c(lz)	c
			$h+k = 2n+1$	−4s(hk)s(lz)	4c(hk)s(lz)	
9	Cc	$I11b$	$k = 2n$	4c(hk)c(lz)	4s(hk)c(lz)	c
			$k = 2n+1$	−4s(hk)s(lz)	4c(hk)s(lz)	
10	$P2/m$	$P12/m1$		4c(hl)c(ky)	0	b
10	$P2/m$	$P112/m$		4c(hk)c(lz)	0	c
11	$P2_1/m$	$P12_1/m1$	$k = 2n$	4c(hl)c(ky)	0	b
			$k = 2n+1$	−4s(hl)s(ky)	0	

Table A1.4.3.3 (*cont.*)

No.	Group symbol		Parity	A	B	Unique axis
	Short	Full				
11	$P2_1/m$	$P112_1/m$	$l = 2n$	4c(hk)c(lz)	0	c
			$l = 2n+1$	−4s(hk)s(lz)	0	
12	$C2/m$	$C12/m1$		8c(hl)c(ky)	0	b
12	$C2/m$	$A12/m1$		8c(hl)c(ky)	0	b
12	$C2/m$	$I12/m1$		8c(hl)c(ky)	0	b
12	$C2/m$	$A112/m$		8c(hk)c(lz)	0	c
12	$C2/m$	$B112/m$		8c(hk)c(lz)	0	c
12	$C2/m$	$I112/m$		8c(hk)c(lz)	0	c
13	$P2/c$	$P12/c1$	$l = 2n$	4c(hl)c(ky)	0	b
			$l = 2n+1$	−4s(hl)s(ky)	0	
13	$P2/c$	$P12/n1$	$h+l = 2n$	4c(hl)c(ky)	0	b
			$h+l = 2n+1$	−4s(hl)s(ky)	0	
13	$P2/c$	$P12/a1$	$h = 2n$	4c(hl)c(ky)	0	b
			$h = 2n+1$	−4s(hl)s(ky)	0	
13	$P2/c$	$P112/a$	$h = 2n$	4c(hk)c(lz)	0	c
			$h = 2n+1$	−4s(hk)s(lz)	0	
13	$P2/c$	$P112/n$	$h+k = 2n$	4c(hk)c(lz)	0	c
			$h+k = 2n+1$	−4s(hk)s(lz)	0	
13	$P2/c$	$P112/b$	$k = 2n$	4c(hk)c(lz)	0	c
			$k = 2n+1$	−4s(hk)s(lz)	0	
14	$P2_1/c$	$P12_1/c1$	$k+l = 2n$	4c(hl)c(ky)	0	b
			$k+l = 2n+1$	−4s(hl)s(ky)	0	
14	$P2_1/c$	$P12_1/n1$	$h+k+l = 2n$	4c(hl)c(ky)	0	b
			$h+k+l = 2n+1$	−4s(hl)s(ky)	0	
14	$P2_1/c$	$P12_1/a1$	$h+k = 2n$	4c(hl)c(ky)	0	b
			$h+k = 2n+1$	−4s(hl)s(ky)	0	
14	$P2_1/c$	$P112_1/a$	$h+l = 2n$	4c(hk)c(lz)	0	c
			$h+l = 2n+1$	−4s(hk)s(lz)	0	
14	$P2_1/c$	$P112_1/n$	$h+k+l = 2n$	4c(hk)c(lz)	0	c
			$h+k+l = 2n+1$	−4s(hk)s(lz)	0	
14	$P2_1/c$	$P112_1/b$	$k+l = 2n$	4c(hk)c(lz)	0	c
			$k+l = 2n+1$	−4s(hk)s(lz)	0	
15	$C2/c$	$C12/c1$	$l = 2n$	8c(hl)c(ky)	0	b
			$l = 2n+1$	−8s(hl)s(ky)	0	
15	$C2/c$	$A12/n1$	$h+l = 2n$	8c(hl)c(ky)	0	b
			$h+l = 2n+1$	−8s(hl)s(ky)	0	
15	$C2/c$	$I12/a1$	$h = 2n$	8c(hl)c(ky)	0	b
			$h = 2n+1$	−8s(hl)s(ky)	0	
15	$C2/c$	$A112/a$	$h = 2n$	8c(hk)c(lz)	0	c
			$h = 2n+1$	−8s(hk)s(lz)	0	
15	$C2/c$	$B112/n$	$h+k = 2n$	8c(hk)c(lz)	0	c
			$h+k = 2n+1$	−8s(hk)s(lz)	0	
15	$C2/c$	$I112/b$	$k = 2n$	8c(hk)c(lz)	0	c
			$k = 2n+1$	−8s(hk)s(lz)	0	

Table A1.4.3.4. *Orthorhombic space groups*

The expressions for A and B for the orthorhombic space groups in their standard settings [as in *IT* A (2005)] contain one, two or four terms of the form

$$\mathrm{pqr} = \mathrm{p}(2\pi hx)\mathrm{q}(2\pi ky)\mathrm{r}(2\pi lz) \qquad \text{(A1.4.3.2)}$$

preceded by a signed numerical constant, where p, q and r can each be either a sine or a cosine function, and the arguments of the functions in any product of the form (A1.4.3.2) are ordered as in (A1.4.3.2). These products are given in this table as ccc, ccs, csc, scc, ssc, scs, css and/or sss, where c and s are abbreviations for 'sin' and 'cos', respectively.

Note that pqr vanishes if at least one of p, q and r is a sine, and the corresponding index h, k or l is zero.

No.	Symbol	Origin	Parity	A	B
16	$P222$			4ccc	−4sss
17	$P222_1$		$l = 2n$	4ccc	−4sss
			$l = 2n + 1$	−4css	4scc
18	$P2_12_12$		$h + k = 2n$	4ccc	−4sss
			$h + k = 2n + 1$	−4ssc	4ccs
19	$P2_12_12_1$		$h + k = 2n$; $k + l = 2n$	4ccc	−4sss
			$h + k = 2n$; $k + l = 2n + 1$	−4css	4scc
			$h + k = 2n + 1$; $k + l = 2n$	−4scs	4csc
			$h + k = 2n + 1$; $k + l = 2n + 1$	−4ssc	4ccs
20	$C222_1$		$l = 2n$	8ccc	−8sss
			$l = 2n + 1$	−8css	8scc
21	$C222$			8ccc	−8sss
22	$F222$			16ccc	−16sss
23	$I222$			8ccc	−8sss
24	$I2_12_12_1$		h, k, l all even	8ccc	−8sss
			$h = 2n$; $k, l = 2n + 1$	−8scs	8csc
			$k = 2n$; $l, h = 2n + 1$	−8ssc	8ccs
			$l = 2n$; $h, k = 2n + 1$	−8css	8scc
25	$Pmm2$			4ccc	4ccs
26	$Pmc2_1$		$l = 2n$	4ccc	4ccs
			$l = 2n + 1$	−4css	4csc
27	$Pcc2$		$l = 2n$	4ccc	4ccs
			$l = 2n + 1$	−4ssc	−4sss
28	$Pma2$		$h = 2n$	4ccc	4ccs
			$h = 2n + 1$	−4ssc	−4sss
29	$Pca2_1$		$h = 2n$; $l = 2n$	4ccc	4ccs
			$h = 2n$; $l = 2n + 1$	−4scs	4scc
			$h = 2n + 1$; $l = 2n$	−4ssc	−4sss
			$h = 2n + 1$; $l = 2n + 1$	−4css	4csc
30	$Pnc2$		$k + l = 2n$	4ccc	4ccs
			$k + l = 2n + 1$	−4ssc	4sss
31	$Pmn2_1$		$h + l = 2n$	4ccc	4ccs
			$h + l = 2n + 1$	−4css	4csc
32	$Pba2$		$h + k = 2n$	4ccc	4ccs
			$h + k = 2n + 1$	−4ssc	−4sss
33	$Pna2_1$		$h + k = 2n$; $l = 2n$	4ccc	4ccs
			$h + k = 2n$; $l = 2n + 1$	−4scs	4scc
			$h + k = 2n + 1$; $l = 2n$	−4ssc	−4sss
			$h + k = 2n + 1$; $l = 2n + 1$	−4css	4csc
34	$Pnn2$		$h + k + l = 2n$	4ccc	4ccs
			$h + k + l = 2n + 1$	−4ssc	−4sss
35	$Cmm2$			8ccc	8ccs
36	$Cmc2_1$		$l = 2n$	8ccc	8ccs
			$l = 2n + 1$	−8css	8csc
37	$Ccc2$		$l = 2n$	8ccc	8ccs
			$l = 2n + 1$	−8ssc	−8sss
38	$Amm2$			8ccc	8ccs
39	$Abm2$		$k = 2n$	8ccc	8ccs
			$k = 2n + 1$	−8ssc	−8sss
40	$Ama2$		$h = 2n$	8ccc	8ccs
			$h = 2n + 1$	−8ssc	−8sss
41	$Aba2$		$h + k = 2n$	8ccc	8ccs
			$h + k = 2n + 1$	−8ssc	−8sss

Table A1.4.3.4 (*cont.*)

No.	Symbol	Origin	Parity	A	B
42	*Fmm*2			16ccc	16ccs
43	*Fdd*2		$h+k+l=4n$	16ccc	16ccs
			$h+k+l=4n+1$	8(ccc − ssc −ccs − sss)	8(ccs − sss + ccc +ssc)
			$h+k+l=4n+2$	−16ssc	−16sss
			$h+k+l=4n+3$	8(ccc − ssc + ccs + sss)	8(ccs − sss − ccc − ssc)
44	*Imm*2			8ccc	8ccs
45	*Iba*2		$l=2n$	8ccc	8ccs
			$l=2n+1$	−8ssc	−8sss
46	*Iam*2		$h=2n$	8ccc	8ccs
			$h=2n+1$	−8ssc	−8sss
47	*Pmmm*			8ccc	0
48	*Pnnn*	(1)	$h+k+l=2n$	8ccc	0
			$h+k+l=2n+1$	0	−8sss
48	*Pnnn*	(2)	$h+k=2n$; $k+l=2n$	8ccc	0
			$h+k=2n$; $k+l=2n+1$	−8ssc	0
			$h+k=2n+1$; $k+l=2n$	−8css	0
			$h+k=2n+1$; $k+l=2n+1$	−8scs	0
49	*Pccm*		$l=2n$	8ccc	0
			$l=2n+1$	−8ssc	0
50	*Pban*	(1)	$h+k=2n$	8ccc	0
			$h+k=2n+1$	0	−8sss
50	*Pban*	(2)	$h=2n$; $k=2n$	8ccc	0
			$h=2n$; $k=2n+1$	−8scs	0
			$h=2n+1$; $k=2n$	−8css	0
			$h=2n+1$; $k=2n+1$	−8ssc	0
51	*Pmma*		$h=2n$	8ccc	0
			$h=2n+1$	−8scs	0
52	*Pnna*		$h=2n$; $k+l=2n$	8ccc	0
			$h=2n$; $k+l=2n+1$	−8ssc	0
			$h=2n+1$; $k+l=2n$	−8css	0
			$h=2n+1$; $k+l=2n+1$	−8scs	0
53	*Pmna*		$h+l=2n$	8ccc	0
			$h+l=2n+1$	−8css	0
54	*Pcca*		$h=2n$; $l=2n$	8ccc	0
			$h=2n$; $l=2n+1$	−8ssc	0
			$h=2n+1$; $l=2n$	−8scs	0
			$h=2n+1$; $l=2n+1$	−8css	0
55	*Pbam*		$h+k=2n$	8ccc	0
			$h+k=2n+1$	−8ssc	0
56	*Pccn*		$h+k=2n$; $h+l=2n$	8ccc	0
			$h+k=2n$; $h+l=2n+1$	−8ssc	0
			$h+k=2n+1$; $h+l=2n$	−8css	0
			$h+k=2n+1$; $h+l=2n+1$	−8scs	0
57	*Pbcm*		$k=2n$; $l=2n$	8ccc	0
			$k=2n$; $l=2n+1$	−8css	0
			$k=2n+1$; $l=2n$	−8ssc	0
			$k=2n+1$; $l=2n+1$	−8scs	0
58	*Pnnm*		$h+k+l=2n$	8ccc	0
			$h+k+l=2n+1$	−8ssc	0
59	*Pmmn*	(1)	$h+k=2n$	8ccc	0
			$h+k=2n+1$	0	8ccs
59	*Pmmn*	(2)	$h=2n$; $k=2n$	8ccc	0
			$h=2n$; $k=2n+1$	−8css	0
			$h=2n+1$; $k=2n$	−8scs	0
			$h=2n+1$; $k=2n+1$	−8ssc	0
60	*Pbcn*		$h+k=2n$; $l=2n$	8ccc	0
			$h+k=2n$; $l=2n+1$	−8css	0
			$h+k=2n+1$; $l=2n$	−8scs	0
			$h+k=2n+1$; $l=2n+1$	−8ssc	0
61	*Pbca*		$h+k=2n$; $k+l=2n$	8ccc	0

Table A1.4.3.4 (*cont.*)

No.	Symbol	Origin	Parity	A	B
			$h+k=2n;\ k+l=2n+1$	-8css	0
			$h+k=2n+1;\ k+l=2n$	-8scs	0
			$h+k=2n+1;\ k+l=2n+1$	-8ssc	0
62	*Pnma*		$h+l=2n;\ k=2n$	8ccc	0
			$h+l=2n;\ k=2n+1$	-8ssc	0
			$h+l=2n+1;\ k=2n$	-8scs	0
			$h+l=2n+1;\ k=2n+1$	-8css	0
63	*Cmcm*		$l=2n$	16ccc	0
			$l=2n+1$	-16css	0
64	*Cmca*		$k+l=2n$	16ccc	0
			$k+l=2n+1$	-16css	0
65	*Cmmm*			16ccc	0
66	*Cccm*		$l=2n$	16ccc	0
			$l=2n+1$	-16ssc	0
67	*Cmma*		$h=2n$	16ccc	0
			$h=2n+1$	-16css	0
68	*Ccca*	(1)	$h+l=2n$	16ccc	0
			$h+l=2n+1$	0	-16sss
68	*Ccca*	(2)	$k=2n;\ l=2n$	16ccc	0
			$k=2n;\ l=2n+1$	-16ssc	0
			$k=2n+1;\ l=2n$	-16scs	0
			$k=2n+1;\ l=2n+1$	-16css	0
69	*Fmmm*			32ccc	0
70	*Fddd*	(1)	$h+k+l=4n$	32ccc	0
			$h+k+l=4n+1$	16(ccc $-$ sss)	A
			$h+k+l=4n+2$	0	-32sss
			$h+k+l=4n+3$	16(ccc $+$ sss)	$-A$
70	*Fddd*	(2)	$h+k=4n;\ k+l=4n;\ l+h=4n$	32ccc	0
			$h+k=4n;\ k+l=4n+2;\ l+h=4n+2$	-32ssc	0
			$h+k=4n+2;\ k+l=4n;\ l+h=4n+2$	-32css	0
			$h+k=4n+2;\ k+l=4n+2;\ l+h=4n$	-32scs	0
			$h+k=4n+2;\ k+l=4n+2;$ $l+h=4n+2$	$-$ 16(ccc $+$ ssc $+$ scs $+$ css)	0
			$h+k=4n+2;\ k+l=4n;\ l+h=4n$	16(ccc $+$ ssc $-$ scs $-$ css)	0
			$h+k=4n;\ k+l=4n+2;\ l+h=4n$	16(ccc $-$ ssc $-$ scs $+$ css)	0
			$h+k=4n;\ k+l=4n;\ l+h=4n+2$	16(ccc $-$ ssc $+$ scs $-$ css)	0
71	*Immm*			16ccc	0
72	*Ibam*		$l=2n$	16ccc	0
			$l=2n+1$	-16ssc	0
73	*Ibca*		$h=2n;\ k=2n$	16ccc	0
			$h=2n;\ k=2n+1$	-16scs	0
			$h=2n+1;\ k=2n$	-16ssc	0
			$h=2n+1;\ k=2n+1$	-16css	0
74	*Imma*		$k=2n$	16ccc	0
			$k=2n+1$	-16css	0

Table A1.4.3.5. *Tetragonal space groups*

The symbols appearing in this table are based on the factorization of the scalar product appearing in equations (1.4.2.19) and (1.4.2.20) into its plane-group and unique-axis components. The symbols are

$$\begin{aligned} \mathrm{P(pq)} &= \mathrm{p}(2\pi hx)\mathrm{q}(2\pi ky) + \mathrm{p}(2\pi hy)\mathrm{q}(2\pi kx) \\ \mathrm{M(pq)} &= \mathrm{p}(2\pi hx)\mathrm{q}(2\pi ky) - \mathrm{p}(2\pi hy)\mathrm{q}(2\pi kx), \end{aligned} \tag{A1.4.3.3}$$

where p and q can each be a sine or a cosine.

Explicit trigonometric functions given in the table follow the convention

$$\mathrm{c}(u) = \cos(2\pi u) \quad \mathrm{s}(u) = \sin(2\pi u).$$

Conditions for vanishing symbols:

$$\begin{gathered} \mathrm{P(ss)} = \mathrm{M(ss)} = 0 \text{ if } h = 0 \text{ or } k = 0, \\ \mathrm{P(sc)} = \mathrm{M(sc)} = 0 \text{ if } h = 0, \\ \mathrm{P(cs)} = \mathrm{M(cs)} = 0 \text{ if } k = 0, \\ \mathrm{M(cc)} = \mathrm{M(ss)} = 0 \text{ if } h = k \text{ or } h = -k, \end{gathered}$$

and any explicit sine function vanishes if all the indices (h and k, or l) appearing in its argument are zero.

$P4$ [No. 75]

hkl	A	B
All	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{s}(lz)$

$P4_1$ [No. 76] (enantiomorphous to $P4_3$ [No. 78])

l	A	B
$4n$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{s}(lz)$
$4n+1$	$-2[\mathrm{s}(hx+ky)\mathrm{s}(lz) - \mathrm{s}(hy-kx)\mathrm{c}(lz)]$	$2[\mathrm{s}(hx+ky)\mathrm{c}(lz) + \mathrm{s}(hy-kx)\mathrm{s}(lz)]$
$4n+2$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{c}(lz)$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{s}(lz)$
$4n+3$	$-2[\mathrm{s}(hx+ky)\mathrm{s}(lz) + \mathrm{s}(hy-kx)\mathrm{c}(lz)]$	$2[\mathrm{s}(hx+ky)\mathrm{c}(lz) - \mathrm{s}(hy-kx)\mathrm{s}(lz)]$

$P4_2$ [No. 77]

l	A	B
$2n$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{s}(lz)$
$2n+1$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{c}(lz)$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{s}(lz)$

$P4_3$ [No. 78] (enantiomorphous to $P4_1$ [No. 76])

l	A	B
$4n$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{s}(lz)$
$4n+1$	$-2[\mathrm{s}(hx+ky)\mathrm{s}(lz) + \mathrm{s}(hy-kx)\mathrm{c}(lz)]$	$2[\mathrm{s}(hx+ky)\mathrm{c}(lz) - \mathrm{s}(hy-kx)\mathrm{s}(lz)]$
$4n+2$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{c}(lz)$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{s}(lz)$
$4n+3$	$-2[\mathrm{s}(hx+ky)\mathrm{s}(lz) - \mathrm{s}(hy-kx)\mathrm{c}(lz)]$	$2[\mathrm{s}(hx+ky)\mathrm{c}(lz) + \mathrm{s}(hy-kx)\mathrm{s}(lz)]$

$I4$ [No. 79]

hkl	A	B
All	$4[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$4[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{s}(lz)$

$I4_1$ [No. 80]

$2h+l$	A	B
$4n$	$4[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$4[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{s}(lz)$
$4n+1$	$4[\mathrm{c}(hx+ky)\mathrm{c}(lz) + \mathrm{c}(hy-kx)\mathrm{s}(lz)]$	$4[\mathrm{c}(hx+ky)\mathrm{s}(lz) - \mathrm{c}(hy-kx)\mathrm{c}(lz)]$
$4n+2$	$4[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{c}(lz)$	$4[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{s}(lz)$
$4n+3$	$4[\mathrm{c}(hx+ky)\mathrm{c}(lz) - \mathrm{c}(hy-kx)\mathrm{s}(lz)]$	$4[\mathrm{c}(hx+ky)\mathrm{s}(lz) + \mathrm{c}(hy-kx)\mathrm{c}(lz)]$

$P\bar{4}$ [No. 81]

hkl	A	B
All	$2[\mathrm{P(cc)} - \mathrm{M(ss)}]\mathrm{c}(lz)$	$2[\mathrm{M(cc)} - \mathrm{P(ss)}]\mathrm{s}(lz)$

Table A1.4.3.5 (*cont.*)

$I\bar{4}$ [No. 82]

hkl	*A*	*B*
All	4[P(cc) − M(ss)]c(lz)	4[M(cc) − P(ss)]s(lz)

$P4/m$ [No. 83]

hkl	*A*	*B*
All	4[P(cc) − M(ss)]c(lz)	0

$P4_2/m$ [No. 84] ($B = 0$ for all h, k, l)

l	*A*
$2n$	4[P(cc) − M(ss)]c(lz)
$2n + 1$	4[M(cc) − P(ss)]c(lz)

$P4/n$ [No. 85, Origin 1]

$h + k$	*A*	*B*
$2n$	4[P(cc) − M(ss)]c(lz)	0
$2n + 1$	0	4[M(cc) − P(ss)]s(lz)

$P4/n$ [No. 85, Origin 2] ($B = 0$ for all h, k, l)

h	*k*	*A*
$2n$	$2n$	4[P(cc) − M(ss)]c(lz)
$2n$	$2n + 1$	−4[P(cs) + M(sc)]s(lz)
$2n + 1$	$2n$	−4[M(cs) + P(sc)]s(lz)
$2n + 1$	$2n + 1$	4[M(cc) − P(ss)]c(lz)

$P4_2/n$ [No. 86, Origin 1]

$h + k + l$	*A*	*B*
$2n$	4[P(cc) − M(ss)]c(lz)	0
$2n + 1$	0	4[M(cc) − P(ss)]s(lz)

$P4_2/n$ [No. 86, Origin 2] ($B = 0$ for all h, k, l)

$h + k$	$k + l$	$h + l$	*A*
$2n$	$2n$	$2n$	4[P(cc) − M(ss)]c(lz)
$2n$	$2n + 1$	$2n + 1$	4[M(cc) − P(ss)]c(lz)
$2n + 1$	$2n + 1$	$2n$	−4[M(cs) + P(sc)]s(lz)
$2n + 1$	$2n$	$2n + 1$	−4[P(cs) + M(sc)]s(lz)

$I4/m$ [No. 87]

hkl	*A*	*B*
All	8[P(cc) − M(ss)]c(lz)	0

$I4_1/a$ [No. 88, Origin 1]

$2k + l$	*A*	*B*
$4n$	8[P(cc) − M(ss)]c(lz)	0
$4n + 1$	4[P(cc) − M(ss)]c(lz) + [M(cc) − P(ss)]s(lz)	*A*
$4n + 2$	0	8[M(cc) − P(ss)]s(lz)
$4n + 3$	4[P(cc) − M(ss)]c(lz) − [M(cc) − P(ss)]s(lz)	−*A*

$I4_1/a$ [No. 88, Origin 2] ($B = 0$ for all h, k, l)

h	*k*	$h + k + l$	*A*
$2n$	$2n$	$4n$	8[P(cc) − M(ss)]c(lz)
$2n$	$2n + 1$	$4n$	−8[s($hx + ky$)s(lz) − c($hy − kx$)c(lz)]
$2n + 1$	$2n$	$4n$	8[c($hx + ky$)c(lz) − s($hy − kx$)s(lz)]
$2n + 1$	$2n + 1$	$4n$	−8[M(cs) + P(sc)]s(lz)
$2n$	$2n$	$4n + 2$	8[M(cc) − P(ss)]c(lz)
$2n$	$2n + 1$	$4n + 2$	−8[s($hx + ky$)s(lz) + c($hy − kx$)c(lz)]
$2n + 1$	$2n$	$4n + 2$	8[c($hx + ky$)c(lz) + s($hy − kx$)s(lz)]
$2n + 1$	$2n + 1$	$4n + 2$	−8[P(cs) + M(sc)]s(lz)

Table A1.4.3.5 (*cont.*)

*P*422 [No. 89]

hkl	*A*	*B*
All	4P(cc)c(lz)	−4M(ss)s(lz)

*P*42$_1$2 [No. 90]

$h+k$	*A*	*B*
$2n$	4P(cc)c(lz)	−4M(ss)s(lz)
$2n+1$	−4P(ss)c(lz)	4M(cc)s(lz)

*P*4$_1$22 [No. 91] (enantiomorphous to *P*4$_3$22 [No. 95])

l	*A*	*B*
$4n$	4P(cc)c(lz)	−4M(ss)s(lz)
$4n+1$	−4[s(hx)c(ky)s(lz) − c(kx)s(hy)c(lz)]	4[c(hx)s(ky)c(lz) − s(kx)c(hy)s(lz)]
$4n+2$	4M(cc)c(lz)	−4P(ss)s(lz)
$4n+3$	−4[s(hx)c(ky)s(lz) + c(kx)s(hy)c(lz)]	4[c(hx)s(ky)c(lz) + s(kx)c(hy)s(lz)]

*P*4$_1$2$_1$2 [No. 92] (enantiomorphous to *P*4$_3$2$_1$2 [No. 96])

$2h+2k+l$	*A*	*B*
$4n$	4P(cc)c(lz)	−4M(ss)s(lz)
$4n+1$	2{[P(sc) − P(cs)]c(lz) − [M(cs) − M(sc)]s(lz)}	2{[P(sc) + P(cs)]c(lz) + [M(cs) − M(sc)]s(lz)}
$4n+2$	−4P(ss)c(lz)	4M(cc)s(lz)
$4n+3$	−2{[P(sc) − P(cs)]c(lz) + [M(cs) + M(sc)]s(lz)}	2{[P(sc) + P(cs)]c(lz) − [M(cs) − M(sc)s(lz)}

*P*4$_2$22 [No. 93]

l	*A*	*B*
$2n$	4P(cc)c(lz)	−4M(ss)s(lz)
$2n+1$	4M(cc)c(lz)	−4P(ss)s(lz)

*P*4$_2$2$_1$2 [No. 94]

$h+k+l$	*A*	*B*
$2n$	4P(cc)c(lz)	−4M(ss)s(lz)
$2n+1$	−4P(ss)c(lz)	4M(cc)s(lz)

*P*4$_3$22 [No. 95] (enantiomorphous to *P*4$_1$22 [No. 91])

l	*A*	*B*
$4n$	4P(cc)c(lz)	−4M(ss)s(lz)
$4n+1$	−4[s(hx)c(ky)s(lz) + c(kx)s(hy)c(lz)]	4[c(hx)s(ky)c(lz) + s(kx)c(hy)c(lz)]
$4n+2$	4M(cc)c(lz)	−4P(ss)s(lz)
$4n+3$	−4[s(hx)c(ky)s(lz) − c(kx)s(hy)c(lz)]	4[c(hx)s(ky)c(lz) − s(kx)c(hy)c(lz)]

*P*4$_3$2$_1$2 [No. 96] (enantiomorphous to *P*4$_1$2$_1$2 [No. 92])

$2h+2k+l$	*A*	*B*
$4n$	4P(cc)c(lz)	−4M(ss)s(lz)
$4n+1$	−2{[P(sc) − P(cs)]c(lz) + [M(cs) + M(sc)]s(lz)}	2{[P(sc) + P(cs)]c(lz) − [M(cs) − M(sc)]s(lz)}
$4n+2$	−4P(ss)c(lz)	4M(cc)s(lz)
$4n+3$	2{[P(sc) − P(cs)]c(lz) − [M(cs) + M(sc)]s(lz)}	2{[P(sc) + P(cs)]c(lz) + [M(cs) − M(sc)]s(lz)}

*I*422 [No. 97]

hkl	*A*	*B*
All	8P(cc)c(lz)	−8M(ss)s(lz)

*I*4$_1$22 [No. 98]

$2k+l$	*A*	*B*
$4n$	8P(cc)c(lz)	−8M(ss)s(lz)
$4n+1$	4{[P(cc) − P(ss)]c(lz) + [M(cc) + M(ss)]s(lz)}	4{[P(cc) + P(ss)]c(lz) + [M(cc) − M(ss)]s(lz)}
$4n+2$	−8P(ss)c(lz)	8M(cc)s(lz)
$4n+3$	4{[P(cc) − P(ss)]c(lz) − [M(cc) + M(ss)]s(lz)}	−4{[P(cc) + P(ss)]c(lz) − [M(cc) − M(ss)]s(lz)}

Table A1.4.3.5 (*cont.*)

*P*4*mm* [No. 99]

hkl	*A*	*B*
All	4P(cc)c(lz)	4P(cc)s(lz)

*P*4*bm* [No. 100]

$h+k$	*A*	*B*
$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	−4M(ss)c(lz)	−4M(ss)s(lz)

*P*4$_2$*cm* [No. 101]

l	*A*	*B*
$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	−4P(ss)c(lz)	−4P(ss)s(lz)

*P*4$_2$*nm* [No. 102]

$h+k+l$	*A*	*B*
$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	−4P(ss)c(lz)	−4P(ss)s(lz)

*P*4*cc* [No. 103]

l	*A*	*B*
$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	−4M(ss)c(lz)	−4M(ss)s(lz)

*P*4*nc* [No. 104]

$h+k+l$	*A*	*B*
$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	−4M(ss)c(lz)	−4M(ss)s(lz)

*P*4$_2$*mc* [No. 105]

l	*A*	*B*
$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	4M(cc)c(lz)	4M(cc)s(lz)

*P*4$_2$*bc* [No. 106]

$h+k$	*l*	*A*	*B*
$2n$	$2n$	4P(cc)c(lz)	4P(cc)s(lz)
$2n+1$	$2n$	−4M(ss)c(lz)	−4M(ss)s(lz)
$2n$	$2n+1$	4M(cc)c(lz)	4M(cc)s(lz)
$2n+1$	$2n+1$	−4P(ss)c(lz)	−4P(ss)s(lz)

*I*4*mm* [No. 107]

hkl	*A*	*B*
All	8P(cc)c(lz)	8P(cc)s(lz)

*I*4*cm* [No. 108]

l	*A*	*B*
$2n$	8P(cc)c(lz)	8P(cc)s(lz)
$2n+1$	−8M(ss)c(lz)	−8M(ss)s(lz)

*I*4$_1$*md* [No. 109]

$2k+l$	*A*	*B*
$4n$	8P(cc)c(lz)	8P(cc)s(lz)
$4n+1$	8[c(hx)c(ky)c(lz) − c(kx)c(hy)s(lz)]	8[c(hx)c(ky)s(lz) + c(kx)c(hy)c(lz)]
$4n+2$	8M(cc)c(lz)	8M(cc)s(lz)
$4n+3$	8[c(hx)c(ky)c(lz) + c(kx)c(hy)s(lz)]	8[c(hx)c(ky)s(lz) − c(kx)c(hy)c(lz)]

Table A1.4.3.5 (*cont.*)

$I4_1cd$ [No. 110]

$2k+l$	A	B
$4n$	8P(cc)c(lz)	8P(cc)s(lz)
$4n+1$	-8[s(hx)s(ky)c(lz) + s(kx)s(hy)s(lz)]	-8[s(hx)s(ky)s(lz) − s(kx)s(hy)c(lz)]
$4n+2$	8M(cc)c(lz)	8M(cc)s(lz)
$4n+3$	-8[s(hx)s(ky)c(lz) − s(kx)s(hy)s(lz)]	-8[s(hx)s(ky)s(lz) + s(kx)s(hy)c(lz)]

$P\bar{4}2m$ [No. 111]

hkl	A	B
All	4P(cc)c(lz)	-4P(ss)s(lz)

$P\bar{4}2c$ [No. 112]

l	A	B
$2n$	4P(cc)c(lz)	-4P(ss)s(lz)
$2n+1$	-4M(ss)c(lz)	4M(cc)s(lz)

$P\bar{4}2_1m$ [No. 113]

$h+k$	A	B
$2n$	4P(cc)c(lz)	-4P(ss)s(lz)
$2n+1$	-4M(ss)c(lz)	4M(cc)s(lz)

$P\bar{4}2_1c$ [No. 114]

$h+k+l$	A	B
$2n$	4P(cc)c(lz)	-4P(ss)s(lz)
$2n+1$	-4M(ss)c(lz)	4M(cc)s(lz)

$P\bar{4}m2$ [No. 115]

hkl	A	B
All	4P(cc)c(lz)	4M(cc)s(lz)

$P\bar{4}c2$ [No. 116]

l	A	B
$2n$	4P(cc)c(lz)	4M(cc)s(lz)
$2n+1$	-4M(ss)c(lz)	-4P(ss)s(lz)

$P\bar{4}b2$ [No. 117]

$h+k$	A	B
$2n$	4P(cc)c(lz)	4M(cc)s(lz)
$2n+1$	-4M(ss)c(lz)	-4P(ss)s(lz)

$P\bar{4}n2$ [No. 118]

$h+k+l$	A	B
$2n$	4P(cc)c(lz)	4M(cc)s(lz)
$2n+1$	-4M(ss)c(lz)	-4P(ss)s(lz)

$I\bar{4}m2$ [No. 119]

hkl	A	B
All	8P(cc)c(lz)	8M(cc)s(lz)

$I\bar{4}c2$ [No. 120]

l	A	B
$2n$	8P(cc)c(lz)	8M(cc)s(lz)
$2n+1$	-8M(ss)c(lz)	-8P(ss)s(lz)

$I\bar{4}2m$ [No. 121]

hkl	A	B
All	8P(cc)c(lz)	-8P(ss)s(lz)

Table A1.4.3.5 (*cont.*)

$I\bar{4}2d$ [No. 122]

$2h+l$	A	B
$4n$	8P(cc)c(lz)	−8P(ss)s(lz)
$4n+1$	4{[P(cc) − M(ss)]c(lz) − [M(cc) + P(ss)]s(lz)}	−4{[P(cc) + M(ss)]c(lz) − [M(cc) − P(ss)]s(lz)}
$4n+2$	−8M(ss)c(lz)	8M(cc)s(lz)
$4n+3$	4{[P(cc) − M(ss)]c(lz) + [M(cc) + P(ss)]s(lz)}	4{[P(cc) + M(ss)]c(lz) + [M(cc) − P(ss)]s(lz)}

$P4/mmm$ [No. 123]

hkl	A	B
All	8P(cc)c(lz)	0

$P4/mcc$ [No. 124] (B = 0 for all h, k, l)

l	A
$2n$	8P(cc)c(lz)
$2n+1$	−8M(ss)c(lz)

$P4/nbm$ [No. 125, Origin 1]

$h+k$	A	B
$2n$	8P(cc)c(lz)	0
$2n+1$	0	−8M(ss)s(lz)

$P4/nbm$ [No. 125, Origin 2] (B = 0 for all h, k, l)

h	k	A
$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n+1$	−8M(sc)s(lz)
$2n+1$	$2n$	−8M(cs)s(lz)
$2n+1$	$2n+1$	−8P(ss)c(lz)

$P4/nnc$ [No. 126, Origin 1]

$h+k+l$	A	B
$2n$	8P(cc)c(lz)	0
$2n+1$	0	−8M(ss)s(lz)

$P4/nnc$ [No. 126, Origin 2] (B = 0 for all h, k, l)

h	k	l	A
$2n$	$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n$	$2n+1$	−8M(ss)c(lz)
$2n$	$2n+1$	$2n$	−8M(sc)s(lz)
$2n$	$2n+1$	$2n+1$	−8P(cs)s(lz)
$2n+1$	$2n$	$2n$	−8M(cs)s(lz)
$2n+1$	$2n$	$2n+1$	−8P(sc)s(lz)
$2n+1$	$2n+1$	$2n$	−8P(ss)c(lz)
$2n+1$	$2n+1$	$2n+1$	8M(cc)c(lz)

$P4/mbm$ [No. 127] (B = 0 for all h, k, l)

$h+k$	A
$2n$	8P(cc)c(lz)
$2n+1$	−8M(ss)c(lz)

$P4/mnc$ [No. 128] (B = 0 for all h, k, l)

$h+k+l$	A
$2n$	8P(cc)c(lz)
$2n+1$	−8M(ss)c(lz)

$P4/nmm$ [No. 129, Origin 1]

$h+k$	A	B
$2n$	8P(cc)c(lz)	0
$2n+1$	0	8M(cc)s(lz)

Table A1.4.3.5 (*cont.*)

*P*4/*nmm* [No. 129, Origin 2] ($B = 0$ for all h, k, l)

h	k	A
$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n + 1$	−8P(cs)s(lz)
$2n + 1$	$2n$	−8P(sc)s(lz)
$2n + 1$	$2n + 1$	−8P(ss)c(lz)

*P*4/*ncc* [No. 130, Origin 1]

$h + k$	l	A	B
$2n$	$2n$	8P(cc)c(lz)	0
$2n$	$2n + 1$	−8M(ss)c(lz)	0
$2n$ +1	$2n$	0	8M(cc)s(lz)
$2n + 1$	$2n + 1$	0	−8P(ss)s(lz)

*P*4/*ncc* [No. 130, Origin 2] ($B = 0$ for all h, k, l)

h	k	l	A
$2n$	$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n$	$2n + 1$	−8M(ss)c(lz)
$2n$	$2n + 1$	$2n$	−8P(cs)s(lz)
$2n$	$2n + 1$	$2n + 1$	−8M(sc)s(lz)
$2n + 1$	$2n$	$2n$	−8P(sc)s(lz)
$2n + 1$	$2n$	$2n + 1$	−8M(cs)s(lz)
$2n + 1$	$2n + 1$	$2n$	−8P(ss)c(lz)
$2n + 1$	$2n + 1$	$2n + 1$	8M(cc)c(lz)

$P4_2/mmc$ [No. 131] ($B = 0$ for all h, k, l)

l	A
$2n$	8P(cc)c(lz)
$2n + 1$	8M(cc)c(lz)

$P4_2/mcm$ [No. 132] ($B = 0$ for all h, k, l)

l	A
$2n$	8P(cc)c(lz)
$2n + 1$	−8P(ss)c(lz)

$P4_2/nbc$ [No. 133, Origin 1]

$h + k + l$	l	A	B
$2n$	$2n$	8P(cc)c(lz)	0
$2n$	$2n + 1$	−8M(ss)c(lz)	0
$2n$ +1	$2n$	0	−8P(ss)s(lz)
$2n + 1$	$2n + 1$	0	8M(cc)s(lz)

$P4_2/nbc$ [No. 133, Origin 2] ($B = 0$ for all h, k, l)

h	k	l	A
$2n$	$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n$	$2n + 1$	8M(cc)c(lz)
$2n$	$2n + 1$	$2n$	−8M(sc)s(lz)
$2n$	$2n + 1$	$2n + 1$	−8P(sc)s(lz)
$2n + 1$	$2n$	$2n$	−8M(cs)s(lz)
$2n + 1$	$2n$	$2n + 1$	−8P(cs)s(lz)
$2n + 1$	$2n + 1$	$2n$	−8P(ss)c(lz)
$2n + 1$	$2n + 1$	$2n + 1$	−8M(ss)c(lz)

$P4_2/nnm$ [No. 134, Origin 1]

$h + k + l$	A	B
$2n$	8P(cc)c(lz)	0
$2n + 1$	0	−8P(ss)s(lz)

Table A1.4.3.5 (*cont.*)

$P4_2/nnm$ [No. 134, Origin 2] ($B = 0$ for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$2n$	$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n+1$	$2n+1$	−8P(ss)c(lz)
$2n+1$	$2n+1$	$2n$	−8M(sc)s(lz)
$2n+1$	$2n$	$2n+1$	−8M(cs)s(lz)

$P4_2/mbc$ [No. 135] ($B = 0$ for all h, k, l)

$h+k$	l	A
$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n+1$	8M(cc)c(lz)
$2n+1$	$2n$	−8M(ss)c(lz)
$2n+1$	$2n+1$	−8P(ss)c(lz)

$P4_2/mnm$ [No. 136] ($B = 0$ for all h, k, l)

$h+k+l$	A
$2n$	8P(cc)c(lz)
$2n+1$	−8P(ss)c(lz)

$P4_2/nmc$ [No. 137, Origin 1]

$h+k+l$	A	B
$2n$	8P(cc)c(lz)	0
$2n+1$	0	8M(cc)s(lz)

$P4_2/nmc$ [No. 137, Origin 2] ($B = 0$ for all h, k, l)

h	k	l	A
$2n$	$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n$	$2n+1$	8M(cc)c(lz)
$2n$	$2n+1$	$2n$	−8P(cs)s(lz)
$2n$	$2n+1$	$2n+1$	−8M(cs)s(lz)
$2n+1$	$2n$	$2n$	−8P(sc)s(lz)
$2n+1$	$2n$	$2n+1$	−8M(sc)s(lz)
$2n+1$	$2n+1$	$2n$	−8P(ss)c(lz)
$2n+1$	$2n+1$	$2n+1$	−8M(ss)c(lz)

$P4_2/ncm$ [No. 138, Origin 1]

$h+k$	l	A	B
$2n$	$2n$	8P(cc)c(lz)	0
$2n+1$	$2n+1$	−8M(ss)c(lz)	0
$2n+1$	$2n$	0	8M(cc)s(lz)
$2n$	$2n+1$	0	−8P(ss)s(lz)

$P4_2/ncm$ [No. 138, Origin 2] ($B = 0$ for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$2n$	$2n$	$2n$	8P(cc)c(lz)
$2n$	$2n+1$	$2n+1$	−8P(ss)c(lz)
$2n+1$	$2n+1$	$2n$	−8P(cs)s(lz)
$2n+1$	$2n$	$2n+1$	−8P(sc)s(lz)

$I4/mmm$ [No. 139]

hkl	A	B
All	16P(cc)c(lz)	0

$I4/mcm$ [No. 140] ($B = 0$ for all h, k, l)

l	A
$2n$	16P(cc)c(lz)
$2n+1$	−16M(ss)c(lz)

Table A1.4.3.5 (*cont.*)

$I4_1/amd$ [No. 141, Origin 1]

$2h+l$	A	B
$4n$	16P(cc)c(lz)	0
$4n+1$	8[P(cc)c(lz) − M(cc)s(lz)]	$-A$
$4n+2$	0	16M(cc)s(lz)
$4n+3$	8[P(cc)c(lz) + M(cc)s(lz)]	A

$I4_1/amd$ [No. 141, Origin 2] (B = 0 for all h, k, l)

h	k	$h+k+l$	A
$2n$	$2n$	$4n$	16P(cc)c(lz)
$2n$	$2n+1$	$4n$	−16[c(hx)s(ky)s(lz) + c(kx)c(hy)c(lz)]
$2n+1$	$2n$	$4n$	16[c(hx)c(ky)c(lz) + c(kx)s(hy)s(lz)]
$2n+1$	$2n+1$	$4n$	−16[c(hx)s(ky)s(lz) + c(kx)s(hy)s(lz)]
$2n$	$2n$	$4n+2$	16M(cc)c(lz)
$2n$	$2n+1$	$4n+2$	−16[c(hx)s(ky)s(lz) − c(kx)c(hy)c(lz)]
$2n+1$	$2n$	$4n+2$	16[c(hx)c(ky)c(lz) − c(kx)s(hy)s(lz)]
$2n+1$	$2n+1$	$4n+2$	−16[c(hx)s(ky)s(lz) − c(kx)s(hy)s(lz)]

$I4_1/acd$ [No. 142, Origin 1]

$2h+l$	A	B
$4n$	16P(cc)c(lz)	0
$4n+1$	−8[M(ss)c(lz) − P(ss)s(lz)]	$-A$
$4n+2$	0	16M(cc)s(lz)
$4n+3$	−8[M(ss)c(lz) + P(ss)s(lz)]	A

$I4_1/acd$ [No. 142, Origin 2] (B = 0 for all h, k, l)

h	k	$h+k+l$	A
$2n$	$2n$	$4n$	16P(cc)c(lz)
$2n$	$2n+1$	$4n$	−16[s(hx)c(ky)s(lz) + s(kx)s(hy)c(lz)]
$2n+1$	$2n$	$4n$	−16[s(hx)s(ky)c(lz) + s(kx)c(hy)s(lz)]
$2n+1$	$2n+1$	$4n$	−16[c(hx)s(ky)s(lz) + c(kx)s(hy)s(lz)]
$2n$	$2n$	$4n+2$	16M(cc)c(lz)
$2n$	$2n+1$	$4n+2$	−16[s(hx)c(ky)s(lz) − s(kx)s(hy)c(lz)]
$2n+1$	$2n$	$4n+2$	−16[s(hx)s(ky)c(lz) − s(kx)c(hy)s(lz)]
$2n+1$	$2n+1$	$4n+2$	−16[c(hx)s(ky)s(lz) − c(kx)s(hy)s(lz)]

Table A1.4.3.6. *Trigonal and hexagonal space groups*

The table lists the expressions for A and B for the space groups belonging to the hexagonal family. For the space groups that are referred to hexagonal axes the expressions are given in terms of symbols related to the decomposition of the scalar products into their plane-group and unique-axis components [*cf.* equations (1.4.3.10)–(1.4.3.12)]. The symbols for the seven rhombohedral space groups in their rhombohedral-axes representation are the same as those used for the cubic space groups [*cf.* equations (1.4.3.4) and (1.4.3.5), and the notes at the start of Table A1.4.3.7]. Factors of the forms $\cos(2\pi x)$ and $\sin(2\pi x)$ are abbreviated by $\mathrm{c}(x)$ and $\mathrm{s}(x)$, respectively. All the symbols used in this table are repeated below. Most expressions are given in terms of

$$\begin{aligned}
\mathrm{C}(hki) &= \mathrm{c}(p_1) + \mathrm{c}(p_2) + \mathrm{c}(p_3),\\
\mathrm{C}(khi) &= \mathrm{c}(q_1) + \mathrm{c}(q_2) + \mathrm{c}(q_3) \text{ and}\\
\mathrm{S}(hki) &= \mathrm{s}(p_1) + \mathrm{s}(p_2) + \mathrm{s}(p_3),\\
\mathrm{S}(khi) &= \mathrm{s}(q_1) + \mathrm{s}(q_2) + \mathrm{s}(q_3),
\end{aligned} \tag{A1.4.3.4}$$

where

$$\begin{aligned}
p_1 &= hx + ky,\ p_2 = kx + iy,\ p_3 = ix + hy,\\
q_1 &= kx + hy,\ q_2 = hx + iy,\ q_3 = ix + ky,
\end{aligned} \tag{A1.4.3.5}$$

and the abbreviations

$$\begin{aligned}
\mathrm{PH(cc)} &= \mathrm{C}(hki) + \mathrm{C}(khi),\\
\mathrm{PH(ss)} &= \mathrm{S}(hki) + \mathrm{S}(khi),\\
\mathrm{MH(cc)} &= \mathrm{C}(hki) - \mathrm{C}(khi) \text{ and}\\
\mathrm{MH(ss)} &= \mathrm{S}(hki) - \mathrm{S}(khi).
\end{aligned} \tag{A1.4.3.6}$$

In addition, the following abbreviations are employed for some space groups:

$$u_1 = lz,\ u_2 = lz + \tfrac{1}{3} \text{ and } u_3 = lz - \tfrac{1}{3}.$$

Conditons for vanishing symbols:

$$\begin{aligned}
&\mathrm{S}(hki) = \mathrm{S}(khi) = 0 \text{ if } h = k = 0,\\
&\mathrm{PH(ss)} = 0 \text{ if } h = -k \text{ (or } k = -i \text{ or } i = -h),\\
&\mathrm{MH(cc)} = 0 \text{ if } |h| = |k| \text{ (or } |k| = |i| \text{ or } |i| = |h|)
\end{aligned}$$

and any explicit sine function vanishes if all the indices (h and k, or l) appearing in its argument are zero.

$P3$ [No. 143]

hkl	*A*	*B*
All	$\mathrm{C}(hki)\mathrm{c}(lz) - \mathrm{S}(hki)\mathrm{s}(lz)$	$\mathrm{C}(hki)\mathrm{s}(lz) + \mathrm{S}(hki)\mathrm{c}(lz)$

$P3_1$ [No. 144] (enantiomorphous to $P3_2$ [No. 145])

l	*A*	*B*
$3n$	as for $P3$ [No. 143]	
$3n+1$	$\mathrm{c}(p_1+u_1) + \mathrm{c}(p_2+u_2) + \mathrm{c}(p_3+u_3)$	$\mathrm{s}(p_1+u_1) + \mathrm{s}(p_2+u_2) + \mathrm{s}(p_3+u_3)$
$3n+2$	$\mathrm{c}(p_1+u_1) + \mathrm{c}(p_2+u_3) + \mathrm{c}(p_3+u_2)$	$\mathrm{s}(p_1+u_1) + \mathrm{s}(p_2+u_3) + \mathrm{s}(p_3+u_2)$

$P3_2$ [No. 145] (enantiomorphous to $P3_1$ [No. 144])

l	*A*, *B*
$3n$	as for $P3$ [No. 143]
$3n+1$	as for $l = 3n+2$ in $P3_1$ [No. 144]
$3n+2$	as for $l = 3n+1$ in $P3_1$ [No. 144]

$R3$ [No. 146] (rhombohedral axes)

hkl	*A*	*B*
All	$\mathrm{c}(hx+ky+lz) + \mathrm{c}(kx+ly+hz) + \mathrm{c}(lx+hy+kz)$	$\mathrm{s}(hx+ky+lz) + \mathrm{s}(kx+ly+hz) + \mathrm{s}(lx+hy+kz)$

$R3$ [No. 146] (hexagonal axes)

hkl	*A*	*B*
All	$3[\mathrm{C}(hki)\mathrm{c}(lz) - \mathrm{S}(hki)\mathrm{s}(lz)]$	$3[\mathrm{C}(hki)\mathrm{s}(lz) + \mathrm{S}(hki)\mathrm{c}(lz)]$

$P\bar{3}$ [No. 147]

hkl	*A*	*B*
All	$2[\mathrm{C}(hki)\mathrm{c}(lz) - \mathrm{S}(hki)\mathrm{s}(lz)]$	0

Table A1.4.3.6 (*cont.*)

$R\bar{3}$ [No. 148] (rhombohedral axes)

hkl	*A*	*B*
All	$2[c(hx+ky+lz)+c(kx+ly+hz)+c(lx+hy+kz)]$	0

$R\bar{3}$ [No. 148] (hexagonal axes)

hkl	*A*	*B*
All	$6[C(hki)c(lz)-S(hki)s(lz)]$	0

*P*312 [No. 149]

hkl	*A*	*B*
All	$PH(cc)c(lz)-PH(ss)s(lz)$	$MH(cc)s(lz)+MH(ss)c(lz)$

*P*321 [No. 150]

hkl	*A*	*B*
All	$PH(cc)c(lz)-MH(ss)s(lz)$	$PH(ss)c(lz)+MH(cc)s(lz)$

$P3_112$ [No. 151] (enantiomorphous to $P3_212$ [No. 153])

l	*A*	*B*
$3n$	as for *P*312 [No. 149]	
$3n+1$	$c(p_1+u_1)+c(p_2+u_2)+c(p_3+u_3)+c(q_1+u_2)+c(q_2+u_3)+c(q_3+u_1)]$	$s(p_1+u_1)+s(p_2+u_2)+s(p_3+u_3)-s(q_1+u_2)-s(q_2+u_3)-s(q_3+u_1)$
$3n+2$	$c(p_1+u_1)+c(p_2+u_3)+c(p_3+u_2)+c(q_1+u_3)+c(q_2+u_2)+c(q_3+u_1)$	$s(p_1+u_1)+s(p_2+u_3)+s(p_3+u_2)-s(q_1+u_3)-s(q_2+u_2)-s(q_3+u_1)$

$P3_121$ [No. 152] (enantiomorphous to $P3_221$ [No. 154])

l	*A*	*B*
$3n$	as for *P*321 [No. 150]	
$3n+1$	$c(p_1+u_1)+c(p_2+u_2)+c(p_3+u_3)+c(q_1-u_1)+c(q_2-u_2)+c(q_3-u_3)$	$s(p_1+u_1)+s(p_2+u_2)+s(p_3+u_3)+s(q_1-u_1)+s(q_2-u_2)+s(q_3-u_3)$
$3n+2$	$c(p_1+u_1)+c(p_2+u_3)+c(p_3+u_2)+c(q_1-u_1)+c(q_2-u_3)+c(q_3-u_2)$	$s(p_1+u_1)+s(p_2+u_3)+s(p_3+u_2)+s(q_1-u_1)+s(q_2-u_3)+s(q_3-u_2)$

$P3_212$ [No. 153] (enantiomorphous to $P3_112$ [No. 151])

l	*A*, *B*
$3n$	as for *P*312 [No. 149]
$3n+1$	as for $l=3n+2$ in $P3_112$ [No. 151]
$3n+2$	as for $l=3n+1$ in $P3_112$ [No. 151]

$P3_221$ [No. 154] (enantiomorphous to $P3_121$ [No. 152])

l	*A*, *B*
$3n$	as for *P*321 [No. 150]
$3n+1$	as for $l=3n+2$ in $P3_121$ [No. 152]
$3n+2$	as for $l=3n+1$ in $P3_121$ [No. 152]

*R*32 [No. 155] (rhombohedral axes)

hkl	*A*	*B*
All	Eccc − Ecss − Escs − Essc + Occc − Ocss − Oscs − Ossc	Escc + Ecsc + Eccs − Esss − Oscc − Ocsc − Occs + Osss

*R*32 [No. 155] (hexagonal axes)

hkl	*A*	*B*
All	$3[PH(cc)c(lz)-MH(ss)s(lz)]$	$3[PH(ss)c(lz)+MH(cc)s(lz)]$

*P*3*m*1 [No. 156]

hkl	*A*	*B*
All	$PH(cc)c(lz)-MH(ss)s(lz)$	$PH(cc)s(lz)+MH(ss)c(lz)$

Table A1.4.3.6 (*cont.*)

*P*31*m* [No. 157]

hkl	*A*	*B*
All	PH(cc)c(lz) − PH(ss)s(lz)	PH(cc)s(lz) + PH(ss)c(lz)

*P*3*c*1 [No. 158]

l	*A*	*B*
$2n$	PH(cc)c(lz) − MH(ss)s(lz)	PH(cc)s(lz) + MH(ss)c(lz)
$2n+1$	MH(cc)c(lz) − PH(ss)s(lz)	PH(ss)c(lz) + MH(cc)s(lz)

*P*31*c* [No. 159]

l	*A*	*B*
$2n$	PH(cc)c(lz) − PH(ss)s(lz)	PH(cc)s(lz) + PH(ss)c(lz)
$2n+1$	MH(cc)c(lz) − MH(ss)s(lz)	MH(cc)s(lz) + MH(ss)c(lz)

*R*3*m* [No. 160] (rhombohedral axes)

hkl	*A*	*B*
All	Eccc − Ecss − Escs − Essc + Occc − Ocss − Oscs − Ossc	Escc + Ecsc + Eccs − Esss + Oscc + Ocsc + Occs − Osss

*R*3*m* [No. 160] (hexagonal axes)

hkl	*A*	*B*
All	3[PH(cc)c(lz) − MH(ss)s(lz)]	3[PH(cc)s(lz) + MH(ss)c(lz)]

*R*3*c* [No. 161] (rhombohedral axes)

$h+k+l$	*A*	*B*
$2n$	Eccc − Ecss − Escs − Essc + Occc − Ocss − Oscs − Ossc	Escc + Ecsc + Eccs − Esss + Oscc + Ocsc + Occs − Osss
$2n+1$	Eccc − Ecss − Escs − Essc − Occc + Ocss + Oscs + Ossc	Escc + Ecsc + Eccs − Esss − Oscc − Ocsc − Occs + Osss

*R*3*c* [No. 161] (hexagonal axes)

l	*A*	*B*
$2n$	3[PH(cc)c(lz) − MH(ss)s(lz)]	3[PH(cc)s(lz) + MH(ss)c(lz)]
$2n+1$	3[MH(cc)c(lz) − PH(ss)s(lz)]	3[PH(ss)c(lz) + MH(cc)s(lz)]

$P\bar{3}1m$ [No. 162] ($B = 0$ for all h, k, l)

A
2[PH(cc)c(lz) − PH(ss)s(lz)]

$P\bar{3}1c$ [No. 163] ($B = 0$ for all h, k, l)

l	*A*
$2n$	2[PH(cc)c(lz) − PH(ss)s(lz)]
$2n+1$	2[MH(cc)c(lz) − MH(ss)s(lz)]

$P\bar{3}m1$ [No. 164] ($B = 0$ for all h, k, l)

A
2[PH(cc)c(lz) − MH(ss)s(lz)]

$P\bar{3}c1$ [No. 165] ($B = 0$ for all h, k, l)

l	*A*
$2n$	2[PH(cc)c(lz) − MH(ss)s(lz)]
$2n+1$	2[MH(cc)c(lz) − PH(ss)s(lz)]

$R\bar{3}m$ [No. 166] (rhombohedral axes, $B = 0$ for all h, k, l)

A
2(Eccc − Ecss − Escs − Essc + Occc − Ocss − Oscs − Ossc)

Table A1.4.3.6 (*cont.*)

$R\bar{3}m$ [No. 166] (hexagonal axes, $B = 0$ for all h, k, l)

A
6[PH(cc)c(lz) − MH(ss)s(lz)]

$R\bar{3}c$ [No. 167] (rhombohedral axes, $B = 0$ for all h, k, l)

$h + k + l$	A
$2n$	2(Eccc − Ecss − Escs − Essc + Occc − Ocss − Oscs − Ossc)
$2n + 1$	2(Eccc − Ecss − Escs − Essc − Occc + Ocss + Oscs + Ossc)

$R\bar{3}c$ [No. 167] (hexagonal axes, $B = 0$ for all h, k, l)

l	A
$2n$	6[PH(cc)c(lz) − MH(ss)s(lz)]
$2n + 1$	6[MH(cc)c(lz) − PH(ss)s(lz)]

$P6$ [No. 168]

hkl	A	B
All	2C(hki)c(lz)	2C(hki)s(lz)

$P6_1$ [No. 169] (enantiomorphous to $P6_5$ [No. 170])

l	A	B
$6n$	as for $P6$ [No.168]	
$6n + 1$	$-2[s(p_1)s(u_1) + s(p_2)s(u_2) + s(p_3)s(u_3)]$	$2[s(p_1)c(u_1) + s(p_2)c(u_2) + s(p_3)c(u_3)]$
$6n + 2$	$2[c(p_1)c(u_1) + c(p_2)c(u_3) + c(p_3)c(u_2)]$	$2[c(p_1)s(u_1) + c(p_2)s(u_3) + c(p_3)s(u_2)]$
$6n + 3$	-2S(hki)s(lz)	2S(hki)c(lz)
$6n + 4$	$2[c(p_1)c(u_1) + c(p_2)c(u_2) + c(p_3)c(u_3)]$	$2[c(p_1)s(u_1) + c(p_2)s(u_2) + c(p_3)s(u_3)]$
$6n + 5$	$-2[s(p_1)s(u_1) + s(p_2)s(u_3) + s(p_3)s(u_2)]$	$2[s(p_1)c(u_1) + s(p_2)c(u_3) + s(p_3)c(u_2)]$

$P6_5$ [No. 170] (enantiomorphous to $P6_1$ [No. 169])

l	A, B
$6n$	as for $P6$ [No. 168]
$6n + 1$	as for $l = 6n + 5$ in $P6_1$ [No. 169]
$6n + 2$	as for $l = 6n + 4$ in $P6_1$ [No. 169]
$6n + 3$	as for $l = 6n + 3$ in $P6_1$ [No. 169]
$6n + 4$	as for $l = 6n + 2$ in $P6_1$ [No. 169]
$6n + 5$	as for $l = 6n + 1$ in $P6_1$ [No. 169]

$P6_2$ [No. 171] (enantiomorphous to $P6_4$ [No. 172])

l	A, B
$3n$	as for $P6$ [No. 168]
$3n + 1$	as for $l = 6n + 2$ in $P6_1$ [No. 169]
$3n + 2$	as for $l = 6n + 4$ in $P6_1$ [No. 169]

$P6_4$ [No. 172] (enantiomorphous to $P6_2$ [No. 171])

l	A, B
$3n$	as for $P6$ [No. 168]
$3n + 1$	as for $l = 6n + 4$ in $P6_1$ [No.169]
$3n + 2$	as for $l = 6n + 2$ in $P6_1$ [No. 169]

$P6_3$ [No. 173]

l	A, B
$2n$	as for $P6$ [No. 168]
$2n + 1$	as for $l = 6n + 3$ in $P6_1$ [No. 169]

$P\bar{6}$ [No. 174]

hkl	A	B
All	2C(hki)c(lz)	2S(hki)c(lz)

Table A1.4.3.6 (*cont.*)

*P*6/*m* [No. 175]

hkl	*A*	*B*
All	4C(*hki*)c(*lz*)	0

$P6_3/m$ [No. 176]

l	*A*	*B*
$2n$	4C(*hki*)c(*lz*)	0
$2n+1$	−4S(*hki*)s(*lz*)	0

*P*622 [No. 177]

hkl	*A*	*B*
All	2PH(cc)c(*lz*)	2MH(cc)s(*lz*)

$P6_122$ [No. 178] (enantiomorphous to $P6_522$ [No. 179])

l	*A*	*B*
$6n$	as for *P*622 [No. 177]	
$6n+1$	$-2[\mathrm{s}(p_1)\mathrm{s}(u_1)+\mathrm{s}(p_2)\mathrm{s}(u_2)+\mathrm{s}(p_3)\mathrm{s}(u_3)-\mathrm{s}(q_1)\mathrm{s}(u_3)-\mathrm{s}(q_2)\mathrm{s}(u_1)-\mathrm{s}(q_3)\mathrm{s}(u_2)]$	$2[\mathrm{s}(p_1)\mathrm{c}(u_1)+\mathrm{s}(p_2)\mathrm{c}(u_2)+\mathrm{s}(p_3)\mathrm{c}(u_3)+\mathrm{s}(q_1)\mathrm{c}(u_3)+\mathrm{s}(q_2)\mathrm{c}(u_1)+\mathrm{s}(q_3)\mathrm{c}(u_2)]$
$6n+2$	$2[\mathrm{c}(p_1)\mathrm{c}(u_1)+\mathrm{c}(p_2)\mathrm{c}(u_3)+\mathrm{c}(p_3)\mathrm{c}(u_2)+\mathrm{c}(q_1)\mathrm{c}(u_2)+\mathrm{c}(q_2)\mathrm{c}(u_1)+\mathrm{c}(q_3)\mathrm{c}(u_3)]$	$2[\mathrm{c}(p_1)\mathrm{s}(u_1)+\mathrm{c}(p_2)\mathrm{s}(u_3)+\mathrm{c}(p_3)\mathrm{s}(u_2)-\mathrm{c}(q_1)\mathrm{s}(u_2)-\mathrm{c}(q_2)\mathrm{s}(u_1)-\mathrm{c}(q_3)\mathrm{s}(u_3)]$
$6n+3$	−2MH(ss)s(*lz*)	2PH(ss)c(*lz*)
$6n+4$	$2[\mathrm{c}(p_1)\mathrm{c}(u_1)+\mathrm{c}(p_2)\mathrm{c}(u_2)+\mathrm{c}(p_3)\mathrm{c}(u_3)+\mathrm{c}(q_1)\mathrm{c}(u_3)+\mathrm{c}(q_2)\mathrm{c}(u_1)+\mathrm{c}(q_3)\mathrm{c}(u_2)]$	$2[\mathrm{c}(p_1)\mathrm{s}(u_1)+\mathrm{c}(p_2)\mathrm{s}(u_2)+\mathrm{c}(p_3)\mathrm{s}(u_3)-\mathrm{c}(q_1)\mathrm{s}(u_3)-\mathrm{c}(q_2)\mathrm{s}(u_1)-\mathrm{c}(q_3)\mathrm{s}(u_2)]$
$6n+5$	$-2[\mathrm{s}(p_1)\mathrm{s}(u_1)+\mathrm{s}(p_2)\mathrm{s}(u_3)+\mathrm{s}(p_3)\mathrm{s}(u_2)-\mathrm{s}(q_1)\mathrm{s}(u_2)-\mathrm{s}(q_2)\mathrm{s}(u_1)-\mathrm{s}(q_3)\mathrm{s}(u_3)]$	$2[\mathrm{s}(p_1)\mathrm{c}(u_1)+\mathrm{s}(p_2)\mathrm{c}(u_3)+\mathrm{s}(p_3)\mathrm{c}(u_2)+\mathrm{s}(q_1)\mathrm{c}(u_2)+\mathrm{s}(q_2)\mathrm{c}(u_1)+\mathrm{s}(q_3)\mathrm{c}(u_3)]$

$P6_522$ [No. 179] (enantiomorphous to $P6_122$ [No. 178])

l	*A*, *B*
$6n$	as for *P*622 [No. 177]
$6n+1$	as for $l = 6n+5$ in $P6_122$ [No. 178]
$6n+2$	as for $l = 6n+4$ in $P6_122$ [No. 178]
$6n+3$	as for $l = 6n+3$ in $P6_122$ [No. 178]
$6n+4$	as for $l = 6n+2$ in $P6_122$ [No. 178]
$6n+5$	as for $l = 6n+1$ in $P6_122$ [No. 178]

$P6_222$ [No. 180] (enantiomorphous to $P6_422$ [No. 181])

l	*A*, *B*
n	as for *P*622 [No. 177]
$3n+1$	as for $l = 6n+2$ in $P6_122$ [No. 178]
$3n+2$	as for $l = 6n+4$ in $P6_122$ [No.178]

$P6_422$ [No. 181] (enantiomorphous to $P6_222$ [No. 180])

l	*A*, *B*
$3n$	as for *P*622 [No. 177]
$3n+1$	as for $l = 6n+4$ in $P6_122$ [No. 178]
$3n+2$	as for $l = 6n+2$ in $P6_122$ [No. 178]

$P6_322$ [No. 182]

l	*A*, *B*
$2n$	as for *P*622 [No. 177]
$2n+1$	as for $l = 6n+3$ in $P6_122$ [No. 178]

*P*6*mm* [No. 183]

hkl	*A*	*B*
All	2PH(cc)c(*lz*)	2PH(cc)s(*lz*)

Table A1.4.3.6 (*cont.*)

P6cc [No. 184]

l	A	B
$2n$	2PH(cc)c(lz)	2PH(cc)s(lz)
$2n+1$	2MH(cc)c(lz)	2MH(cc)s(lz)

P6$_3$*cm* [No. 185]

l	A	B
$2n$	2PH(cc)c(lz)	2PH(cc)s(lz)
$2n+1$	−2PH(ss)s(lz)	2PH(ss)c(lz)

P6$_3$*mc* [No. 186]

l	A	B
$2n$	2PH(cc)c(lz)	2PH(cc)s(lz)
$2n+1$	−2MH(ss)s(lz)	2MH(ss)c(lz)

P$\bar{6}$*m*2 [No. 187]

hkl	A	B
All	2PH(cc)c(lz)	2MH(ss)c(lz)

P$\bar{6}$*c*2 [No. 188]

l	A	B
$2n$	2PH(cc)c(lz)	2MH(ss)c(lz)
$2n+1$	−2PH(ss)s(lz)	2MH(cc)s(lz)

P$\bar{6}$2*m* [No. 189]

hkl	A	B
All	2PH(cc)c(lz)	2PH(ss)c(lz)

P$\bar{6}$2*c* [No. 190]

l	A	B
$2n$	2PH(cc)c(lz)	2PH(ss)c(lz)
$2n+1$	−2MH(ss)s(lz)	2MH(cc)s(lz)

P6/*mmm* [No. 191]

hkl	A	B
All	4PH(cc)c(lz)	0

P6/*mcc* [No. 192] ($B = 0$ for all h, k, l)

l	A
$2n$	4PH(cc)c(lz)
$2n+1$	4MH(cc)c(lz)

P6$_3$/*mcm* [No. 193] ($B = 0$ for all h, k, l)

l	A
$2n$	4PH(cc)c(lz)
$2n+1$	−4PH(ss)s(lz)

P6$_3$/*mmc* [No. 194] ($B = 0$ for all h, k, l)

l	A
$2n$	4PH(cc)c(lz)
$2n+1$	−4MH(ss)s(lz)

Table A1.4.3.7. *Cubic space groups*

The symbols appearing in this table are related to the pqr representation used with the orthorhombic space groups as follows: Each of the symbols defined below is a sum of three pqr terms, where the order of *hkl* is fixed in each of the three terms and that of *xyz* is permuted.

This table and parts of Table A1.4.3.6 (rhombohedral space groups referred to rhombohedral axes) are given in terms of the following two symbols:

$$\mathrm{Epqr} = \mathrm{p}(hx)\mathrm{q}(ky)\mathrm{r}(lz) + \mathrm{p}(hy)\mathrm{q}(kz)\mathrm{r}(lx) + \mathrm{p}(hz)\mathrm{q}(kx)\mathrm{r}(ly) \tag{A1.4.3.7}$$

and

$$\mathrm{Opqr} = \mathrm{p}(hx)\mathrm{q}(kz)\mathrm{r}(ly) + \mathrm{p}(hz)\mathrm{q}(ky)\mathrm{r}(lx) + \mathrm{p}(hy)\mathrm{q}(kx)\mathrm{r}(lz), \tag{A1.4.3.8}$$

where p, q and r can each be a sine or a cosine, and the factor 2π has been absorbed in the abbreviations (see text). As in Tables A1.4.3.1–A1.4.3.6, cosine and sine are abbreviated by c and s, respectively. The permutation of the coordinates is even in Epqr and odd in Opqr.

Conditions for vanishing symbols:

Epqr = Opqr = 0 if at least one of p, q, r is a sine and the index h, k or l in its argument is zero,

$$\mathrm{Eccc} - \mathrm{Occc} = 0 \text{ if } |h| = |k| \text{ (or } |k| = |l| \text{ or } |l| = |h|),$$
$$\mathrm{Esss} - \mathrm{Osss} = 0 \text{ if } |h| = |k| \text{ (or } |k| = |l| \text{ or } |l| = |h|),$$
$$\mathrm{Ecss} - \mathrm{Ocss} = \mathrm{Escc} - \mathrm{Oscc} = 0 \text{ if } |k| = |l|,$$
$$\mathrm{Escs} - \mathrm{Oscs} = \mathrm{Ecsc} - \mathrm{Ocsc} = 0 \text{ if } |l| = |h| \text{ and}$$
$$\mathrm{Essc} - \mathrm{Ossc} = \mathrm{Eccs} - \mathrm{Occs} = 0 \text{ if } |h| = |k|.$$

*P*23 [No. 195]

hkl	*A*	*B*
All	4Eccc	−4Esss

*F*23 [No. 196]

hkl	*A*	*B*
All	16Eccc	−16Esss

*I*23 [No. 197]

hkl	*A*	*B*
All	8Eccc	−8Esss

$P2_13$ [No. 198]

$h+k$	$k+l$	$h+l$	*A*	*B*
$2n$	$2n$	$2n$	4Eccc	−4Esss
$2n$	$2n+1$	$2n+1$	−4Ecss	4Escc
$2n+1$	$2n$	$2n+1$	−4Escs	4Ecsc
$2n+1$	$2n+1$	$2n$	−4Essc	4Eccs

$I2_13$ [No. 199]

$h+k$	$k+l$	$h+l$	*A*	*B*
$2n$	$2n$	$2n$	8Eccc	−8Esss
$2n+1$	$2n$	$2n+1$	−8Escs	8Ecsc
$2n+1$	$2n+1$	$2n$	−8Essc	8Eccs
$2n$	$2n+1$	$2n+1$	−8Ecss	8Escc

$Pm\bar{3}$ [No. 200]

hkl	*A*	*B*
All	8Eccc	0

$Pn\bar{3}$ (Origin 1) [No. 201]

$h+k+l$	*A*	*B*
$2n$	8Eccc	0
$2n+1$	0	−8Esss

Table A1.4.3.7 (*cont.*)

Pn$\bar{3}$ (Origin 2) [No. 201] (B = 0 for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$2n$	$2n$	$2n$	8Eccc
$2n$	$2n+1$	$2n+1$	−8Essc
$2n+1$	$2n$	$2n+1$	−8Ecss
$2n+1$	$2n+1$	$2n$	−8Escs

Fm$\bar{3}$ [No. 202]

hkl	A	B
All	32Eccc	0

Fd$\bar{3}$ (Origin 1) [No. 203]

$h+k+l$	A	B
$4n$	32Eccc	0
$4n+1$	16(Eccc − Esss)	A
$4n+2$	0	−32Esss
$4n+3$	16(Eccc + Esss)	$-A$

Fd$\bar{3}$ (Origin 2) [No. 203] (B = 0 for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$4n$	$4n$	$4n$	32Eccc
$4n$	$4n+2$	$4n+2$	−32Essc
$4n+2$	$4n$	$4n+2$	−32Ecss
$4n+2$	$4n+2$	$4n$	−32Escs
$4n+2$	$4n+2$	$4n+2$	−16(Eccc + Ecss + Escs + Essc)
$4n+2$	$4n$	$4n$	16(Eccc − Ecss − Escs + Essc)
$4n$	$4n+2$	$4n$	16(Eccc + Ecss − Escs − Essc)
$4n$	$4n$	$4n+2$	16(Eccc − Ecss + Escs −Essc)

Im$\bar{3}$ [No. 204]

hkl	A	B
All	16Eccc	0

Pa$\bar{3}$ [No. 205] (B = 0 for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$2n$	$2n$	$2n$	8Eccc
$2n$	$2n+1$	$2n+1$	−8Ecss
$2n+1$	$2n$	$2n+1$	−8Escs
$2n+1$	$2n+1$	$2n$	−8Essc

Ia$\bar{3}$ [No. 206] (B = 0 for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$2n$	$2n$	$2n$	16Eccc
$2n$	$2n+1$	$2n+1$	−16Ecss
$2n+1$	$2n$	$2n+1$	−16Escs
$2n+1$	$2n+1$	$2n$	−16Essc

*P*432 [No. 207]

hkl	A	B
All	4(Eccc + Occc)	−4(Esss − Osss)

$P4_232$ [No. 208]

$h+k+l$	A	B
$2n$	4(Eccc + Occc)	−4(Esss − Osss)
$2n+1$	4(Eccc − Occc)	−4(Esss + Osss)

Table A1.4.3.7 (*cont.*)

*F*432 [No. 209]

hkl	*A*	*B*
All	16(Eccc + Occc)	−16(Esss − Osss)

$F4_132$ [No. 210]

$h+k+l$	*A*	*B*
$4n$	16(Eccc + Occc)	−16(Esss − Osss)
$4n+1$	16(Eccc − Osss)	−16(Esss − Occc)
$4n+2$	16(Eccc − Occc)	−16(Esss + Osss)
$4n+3$	16(Eccc + Osss)	−16(Esss + Occc)

*I*432 [No. 211]

hkl	*A*	*B*
All	8(Eccc + Occc)	−8(Esss − Osss)

$P4_332$ [No. 212] (enantiomorphous to $P4_132$ [No. 213])

$h+k$	$k+l$	$h+l$	$h+k+l$	*A*	*B*
$2n$	$2n$	$2n$	$4n$	4(Eccc + Occc)	−4(Esss − Osss)
$2n$	$2n+1$	$2n+1$	$4n$	−4(Ecss + Oscs)	4(Escc − Ocsc)
$2n+1$	$2n$	$2n+1$	$4n$	−4(Escs + Ossc)	4(Ecsc − Occs)
$2n+1$	$2n+1$	$2n$	$4n$	−4(Essc + Ocss)	4(Eccs − Oscc)
$2n$	$2n$	$2n$	$4n+1$	4(Eccc − Osss)	−4(Esss − Occc)
$2n$	$2n+1$	$2n+1$	$4n+1$	−4(Ecss − Ocsc)	4(Escc − Oscs)
$2n+1$	$2n$	$2n+1$	$4n+1$	−4(Escs − Occs)	4(Ecsc − Ossc)
$2n+1$	$2n+1$	$2n$	$4n+1$	−4(Essc − Oscc)	4(Eccs − Ocss)
$2n$	$2n$	$2n$	$4n+2$	4(Eccc − Occc)	−4(Esss + Osss)
$2n$	$2n+1$	$2n+1$	$4n+2$	−4(Ecss − Oscs)	4(Escc + Ocsc)
$2n+1$	$2n$	$2n+1$	$4n+2$	−4(Escs − Ossc)	4(Ecsc + Occs)
$2n+1$	$2n+1$	$2n$	$4n+2$	−4(Essc − Ocss)	4(Eccs + Oscc)
$2n$	$2n$	$2n$	$4n+3$	4(Eccc + Osss)	−4(Esss + Occc)
$2n$	$2n+1$	$2n+1$	$4n+3$	−4(Ecss + Ocsc)	4(Escc + Oscs)
$2n+1$	$2n$	$2n+1$	$4n+3$	−4(Escs + Occs)	4(Ecsc + Ossc)
$2n+1$	$2n+1$	$2n$	$4n+3$	−4(Essc + Oscc)	4(Eccs + Ocss)

$P4_132$ [No. 213] (enantiomorphous to $P4_332$ [No. 212])

h	*k*	*l*	$h+k+l$	*A*	*B*
$2n$	$2n$	$2n$	$4n$	4(Eccc + Occc)	−4(Esss − Osss)
$2n$	$2n+1$	$2n+1$	$4n$	−4(Escs + Ossc)	4(Ecsc − Occs)
$2n+1$	$2n$	$2n+1$	$4n$	−4(Essc + Ocss)	4(Eccs − Oscc)
$2n+1$	$2n+1$	$2n$	$4n$	−4(Ecss + Oscs)	4(Escc − Ocsc)
$2n+1$	$2n+1$	$2n+1$	$4n+1$	4(Eccc + Osss)	−4(Esss + Occc)
$2n$	$2n$	$2n+1$	$4n+1$	−4(Ecss + Ocsc)	4(Escc + Oscs)
$2n+1$	$2n$	$2n$	$4n+1$	−4(Escs + Occs)	4(Ecsc + Ossc)
$2n$	$2n+1$	$2n$	$4n+1$	−4(Essc + Oscc)	4(Eccs + Ocss)
$2n$	$2n$	$2n$	$4n+2$	4(Eccc − Occc)	−4(Esss + Osss)
$2n$	$2n+1$	$2n+1$	$4n+2$	−4(Escs − Ossc)	4(Ecsc + Occs)
$2n+1$	$2n$	$2n+1$	$4n+2$	−4(Essc − Ocss)	4(Eccs + Oscc)
$2n+1$	$2n+1$	$2n$	$4n+2$	−4(Ecss − Oscs)	4(Escc + Ocsc)
$2n+1$	$2n+1$	$2n+1$	$4n+3$	4(Eccc − Osss)	−4(Esss − Occc)
$2n$	$2n$	$2n+1$	$4n+3$	−4(Ecss − Ocsc)	4(Escc − Oscs)
$2n+1$	$2n$	$2n$	$4n+3$	−4(Escs − Occs)	4(Ecsc − Ossc)
$2n$	$2n+1$	$2n$	$4n+3$	−4(Essc − Oscc)	4(Eccs − Ocss)

$I4_132$ [No. 214]

h	*k*	*l*	$h+k+l$	*A*	*B*
$2n$	$2n$	$2n$	$4n$	8(Eccc + Occc)	−8(Esss − Osss)
$2n$	$2n+1$	$2n+1$	$4n$	−8(Escs + Ossc)	8(Ecsc − Occs)
$2n+1$	$2n$	$2n+1$	$4n$	−8(Essc + Ocss)	8(Eccs − Oscc)
$2n+1$	$2n+1$	$2n$	$4n$	−8(Ecss + Oscs)	8(Escc − Ocsc)

Table A1.4.3.7 (*cont.*)

h	k	l	$h+k+l$	A	B
$2n$	$2n$	$2n$	$4n+2$	8(Eccc − Occc)	−8(Esss + Osss)
$2n$	$2n+1$	$2n+1$	$4n+2$	−8(Escs − Ossc)	8(Ecsc + Occs)
$2n+1$	$2n$	$2n+1$	$4n+2$	−8(Essc − Ocss)	8(Eccs + Oscc)
$2n+1$	$2n+1$	$2n$	$4n+2$	−8(Ecss − Oscs)	8(Escc + Ocsc)

$P\bar{4}3m$ [No. 215]

hkl	A	B
All	4(Eccc + Occc)	−4(Esss + Osss)

$F\bar{4}3m$ [No. 216]

hkl	A	B
All	16(Eccc + Occc)	−16(Esss + Osss)

$I\bar{4}3m$ [No. 217]

hkl	A	B
All	8(Eccc + Occc)	−8(Esss +Osss)

$P\bar{4}3n$ [No. 218]

$h+k+l$	A	B
$2n$	4(Eccc + Occc)	−4(Esss + Osss)
$2n+1$	4(Eccc − Occc)	−4(Esss − Osss)

$F\bar{4}3c$ [No. 219]

$h+k+l$	A	B
$2n$	16(Eccc + Occc)	−16(Esss + Osss)
$2n+1$	16(Eccc − Occc)	−16(Esss − Osss)

$I\bar{4}3d$ [No. 220]

h	k	l	$h+k+l$	A	B
$2n$	$2n$	$2n$	$4n$	8(Eccc + Occc)	−8(Esss + Osss)
$2n$	$2n+1$	$2n+1$	$4n$	−8(Escs + Ossc)	8(Ecsc + Occs)
$2n+1$	$2n$	$2n+1$	$4n$	−8(Essc + Ocss)	8(Eccs + Oscc)
$2n+1$	$2n+1$	$2n$	$4n$	−8(Ecss + Oscs)	8(Escc + Ocsc)
$2n$	$2n$	$2n$	$4n+2$	8(Eccc − Occc)	−8(Esss − Osss)
$2n$	$2n+1$	$2n+1$	$4n+2$	−8(Escs − Ossc)	8(Ecsc − Occs)
$2n+1$	$2n$	$2n+1$	$4n+2$	−8(Essc − Ocss)	8(Eccs − Oscc)
$2n+1$	$2n+1$	$2n$	$4n+2$	−8(Ecss − Oscs)	8(Escc −Ocsc)

$Pm\bar{3}m$ [No. 221]

hkl	A	B
All	8(Eccc + Occc)	0

$Pn\bar{3}n$ (Origin 1) [No. 222]

$h+k+l$	A	B
$2n$	8(Eccc + Occc)	0
$2n+1$	0	−8(Esss − Osss)

$Pn\bar{3}n$ (Origin 2) [No. 222] (B = 0 for all h, k, l)

h	k	l	A
$2n$	$2n$	$2n$	8(Eccc + Occc)
$2n$	$2n+1$	$2n+1$	−8(Ecss + Ocss)
$2n+1$	$2n$	$2n+1$	−8(Escs + Oscs)
$2n+1$	$2n+1$	$2n$	−8(Essc + Ossc)
$2n+1$	$2n+1$	$2n+1$	8(Eccc − Occc)
$2n+1$	$2n$	$2n$	−8(Ecss − Ocss)
$2n$	$2n+1$	$2n$	−8(Escs − Oscs)
$2n$	$2n$	$2n+1$	−8(Essc − Ossc)

Table A1.4.3.7 (*cont.*)

$Pm\bar{3}n$ [No. 223] ($B = 0$ for all h, k, l)

$h+k+l$	A
$2n$	8(Eccc + Occc)
$2n+1$	8(Eccc − Occc)

$Pn\bar{3}m$ (Origin 1) [No. 224]

$h+k+l$	A	B
$2n$	8(Eccc + Occc)	0
$2n+1$	0	−8(Esss + Osss)

$Pn\bar{3}m$ (Origin 2) [No. 224] ($B = 0$ for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$2n$	$2n$	$2n$	8(Eccc + Occc)
$2n$	$2n+1$	$2n+1$	−8(Essc + Ossc)
$2n+1$	$2n$	$2n+1$	−8(Ecss + Ocss)
$2n+1$	$2n+1$	$2n$	−8(Escs + Oscs)

$Fm\bar{3}m$ [No. 225]

hkl	A	B
All	32(Eccc + Occc)	0

$Fm\bar{3}c$ [No. 226] ($B = 0$ for all h, k, l)

$h+k+l$	A
$2n$	32(Eccc + Occc)
$2n+1$	32(Eccc − Occc)

$Fd\bar{3}m$ (Origin 1) [No. 227]

$h+k+l$	A	B
$4n$	32(Eccc + Occc)	0
$4n+1$	16(Eccc − Esss + Occc − Osss)	A
$4n+2$	0	−32(Esss + Osss)
$4n+3$	16(Eccc + Esss + Occc + Osss)	$-A$

$Fd\bar{3}m$ (Origin 2) [No. 227] ($B = 0$ for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$4n$	$4n$	$4n$	32(Eccc + Occc)
$4n$	$4n+2$	$4n+2$	−32(Essc + Ossc)
$4n+2$	$4n$	$4n+2$	−32(Ecss + Ocss)
$4n+2$	$4n+2$	$4n$	−32(Escs + Oscs)
$4n+2$	$4n+2$	$4n+2$	−16(Eccc + Ecss + Escs + Essc + Occc + Ocss + Oscs + Ossc)
$4n+2$	$4n$	$4n$	16(Eccc − Ecss − Escs + Essc + Occc − Ocss − Oscs + Ossc)
$4n$	$4n+2$	$4n$	16(Eccc + Ecss − Escs − Essc + Occc + Ocss − Oscs − Ossc)
$4n$	$4n$	$4n+2$	16(Eccc − Ecss + Escs − Essc + Occc − Ocss + Oscs − Ossc)

$Fd\bar{3}c$ (Origin 1) [No. 228]

$h+k+l$	A	B
$4n$	32(Eccc + Occc)	0
$4n+1$	16(Eccc + Esss − Occc − Osss)	$-A$
$4n+2$	0	−32(Esss + Osss)
$4n+3$	16(Eccc − Esss − Occc + Osss)	A

$Fd\bar{3}c$ (Origin 2) [No. 228] ($B = 0$ for all h, k, l)

$h+k$	$k+l$	$h+l$	A
$4n$	$4n$	$4n$	32(Eccc + Occc)
$4n$	$4n+2$	$4n+2$	−32(Essc + Ossc)
$4n+2$	$4n$	$4n+2$	−32(Ecss + Ocss)
$4n+2$	$4n+2$	$4n$	−32(Escs + Oscs)
$4n+2$	$4n+2$	$4n+2$	−16(Eccc + Ecss + Escs + Essc − Occc − Ocss − Oscs −Ossc)

Table A1.4.3.7 (*cont.*)

$h+k$	$k+l$	$h+l$	A
$4n+2$	$4n$	$4n$	16(Eccc − Ecss − Escs + Essc − Occc + Ocss + Oscs − Ossc)
$4n$	$4n+2$	$4n$	16(Eccc + Ecss − Escs − Essc − Occc − Ocss + Oscs + Ossc)
$4n$	$4n$	$4n+2$	16(Eccc − Ecss + Escs − Essc −Occc + Ocss − Oscs + Ossc)

$Im\bar{3}m$ [No. 229]

hkl	A	B
All	16(Eccc + Occc)	0

$Ia\bar{3}d$ [No. 230] (B = 0 for all h, k, l)

h	k	l	$h+k+l$	A
$2n$	$2n$	$2n$	$4n$	16(Eccc + Occc)
$2n$	$2n+1$	$2n+1$	$4n$	−16(Escs + Ossc)
$2n+1$	$2n$	$2n+1$	$4n$	−16(Essc + Ocss)
$2n+1$	$2n+1$	$2n$	$4n$	−16(Ecss + Oscs)
$2n$	$2n$	$2n$	$4n+2$	16(Eccc − Occc)
$2n$	$2n+1$	$2n+1$	$4n+2$	−16(Escs − Ossc)
$2n+1$	$2n$	$2n+1$	$4n+2$	−16(Essc − Ocss)
$2n+1$	$2n+1$	$2n$	$4n+2$	−16(Ecss − Oscs)

APPENDIX A1.4.4
Crystallographic space groups in reciprocal space

Table A1.4.4.1. *Crystallographic space groups in reciprocal space*

The table entries are described in detail in Section 1.4.4.1. The general format of an entry is

$$(n)\,\mathbf{h}^T\boldsymbol{P}_n : -\mathbf{h}^T\mathbf{t}_n$$

or

$$(n)\,\mathbf{h}^T\boldsymbol{P}_n :,$$

according as the phase-shift part of the entry is nonzero or zero modolo 2π, respectively.

Notes:

(1) For centrosymmetric space groups with the centre located at the unit-cell origin only those entries are given which correspond to symmetry operations not related by inversion. If the origin in such space groups is chosen elsewhere, all the entries corresponding to the operations of the point group are presented.

(2) For trigonal and hexagonal space groups referred to hexagonal axes the Miller–Bravais indices *hkil* are employed, and for the rhombohedral space groups referred to rhombohedral axes the indices are denoted by *hkl* (*cf. IT* I, 1952).

Point group: 1 Triclinic Laue group: $\bar{1}$

*P*1 No. 1 (1)
(1) *hkl*:

Point group: $\bar{1}$ Triclinic Laue group: $\bar{1}$

P$\bar{1}$ No. 2 (2)
(1) *hkl*:

Point group: 2 Monoclinic Laue group: 2/*m*

*P*2 *P*121 Unique axis *b* No. 3 (3)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

*P*2 *P*112 Unique axis *c* No. 3 (4)
(1) *hkl*: (2) $\overline{hk}l$:

$P2_1$ $P12_11$ Unique axis *b* No. 4 (5)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −010/2

$P2_1$ $P112_1$ Unique axis *c* No. 4 (6)
(1) *hkl*: (2) $\overline{hk}l$: −001/2

*C*2 *C*121 Unique axis *b* No. 5 (7)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

*C*2 *A*121 Unique axis *b* No. 5 (8)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

*C*2 *I*121 Unique axis *b* No. 5 (9)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

*C*2 *A*112 Unique axis *c* No. 5 (10)
(1) *hkl*: (2) $\overline{hk}l$:

*C*2 *B*112 Unique axis *c* No. 5 (11)
(1) *hkl*: (2) $\overline{hk}l$:

*C*2 *I*112 Unique axis *c* No. 5 (12)
(1) *hkl*: (2) $\overline{hk}l$:

Point group: *m* Monoclinic Laue group: 2/*m*

Pm *P*1*m*1 Unique axis *b* No. 6 (13)
(1) *hkl*: (2) $h\bar{k}l$:

Pm *P*11*m* Unique axis *c* No. 6 (14)
(1) *hkl*: (2) $hk\bar{l}$:

Pc *P*1*c*1 Unique axis *b* No. 7 (15)
(1) *hkl*: (2) $h\bar{k}l$: −001/2

Pc *P*1*n*1 Unique axis *b* No. 7 (16)
(1) *hkl*: (2) $h\bar{k}l$: −101/2

Pc *P*1*a*1 Unique axis *b* No. 7 (17)
(1) *hkl*: (2) $h\bar{k}l$: −100/2

Pc *P*11*a* Unique axis *c* No. 7 (18)
(1) *hkl*: (2) $hk\bar{l}$: −100/2

Pc *P*11*n* Unique axis *c* No. 7 (19)
(1) *hkl*: (2) $hk\bar{l}$: −110/2

Pc *P*11*b* Unique axis *c* No. 7 (20)
(1) *hkl*: (2) $hk\bar{l}$: −010/2

Cm *C*1*m*1 Unique axis *b* No. 8 (21)
(1) *hkl*: (2) $h\bar{k}l$:

Cm *A*1*m*1 Unique axis *b* No. 8 (22)
(1) *hkl*: (2) $h\bar{k}l$:

Cm *I*1*m*1 Unique axis *b* No. 8 (23)
(1) *hkl*: (2) $h\bar{k}l$:

Cm *A*11*m* Unique axis *c* No. 8 (24)
(1) *hkl*: (2) $hk\bar{l}$:

Cm *B*11*m* Unique axis *c* No. 8 (25)
(1) *hkl*: (2) $hk\bar{l}$:

Cm *I*11*m* Unique axis *c* No. 8 (26)
(1) *hkl*: (2) $hk\bar{l}$:

Cc *C*1*c*1 Unique axis *b* No. 9 (27)
(1) *hkl*: (2) $h\bar{k}l$: −001/2

Cc *A*1*n*1 Unique axis *b* No. 9 (28)
(1) *hkl*: (2) $h\bar{k}l$: −101/2

Table A1.4.4.1 (*cont.*)

Cc *I*1*a*1 Unique axis *b* No. 9 (29)
(1) *hkl*: (2) $h\bar{k}l$: −100/2

Cc *A*11*a* Unique axis *c* No. 9 (30)
(1) *hkl*: (2) $hk\bar{l}$: −100/2

Cc *B*11*n* Unique axis *c* No. 9 (31)
(1) *hkl*: (2) $hk\bar{l}$: −110/2

Cc *I*11*b* Unique axis *c* No. 9 (32)
(1) *hkl*: (2) $hk\bar{l}$: −010/2

Point group: 2/*m* Monoclinic Laue group: 2/*m*

$P2/m$ $P12/m1$ Unique axis *b* No. 10 (33)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

$P2/m$ $P112/m$ Unique axis *c* No. 10 (34)
(1) *hkl*: (2) $\overline{hk}l$:

$P2_1/m$ $P12_1/m1$ Unique axis *b* No. 11 (35)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −010/2

$P2_1/m$ $P112_1/m$ Unique axis *c* No. 11 (36)
(1) *hkl*: (2) $\overline{hk}l$: −001/2

$C2/m$ $C12/m1$ Unique axis *b* No. 12 (37)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

$C2/m$ $A12/m1$ Unique axis *b* No. 12 (38)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

$C2/m$ $I12/m1$ Unique axis *b* No. 12 (39)
(1) *hkl*: (2) $\bar{h}k\bar{l}$:

$C2/m$ $A112/m$ Unique axis *c* No. 12 (40)
(1) *hkl*: (2) $\overline{hk}l$:

$C2/m$ $B112/m$ Unique axis *c* No. 12 (41)
(1) *hkl*: (2) $\overline{hk}l$:

$C2/m$ $I112/m$ Unique axis *c* No. 12 (42)
(1) *hkl*: (2) $\overline{hk}l$:

$P2/c$ $P12/c1$ Unique axis *b* No. 13 (43)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −001/2

$P2/c$ $P12/n1$ Unique axis *b* No. 13 (44)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −101/2

$P2/c$ $P12/a1$ Unique axis *b* No. 13 (45)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −100/2

$P2/c$ $P112/a$ Unique axis *c* No. 13 (46)
(1) *hkl*: (2) $\overline{hk}l$: −100/2

$P2/c$ $P112/n$ Unique axis *c* No. 13 (47)
(1) *hkl*: (2) $\overline{hk}l$: −110/2

$P2/c$ $P112/b$ Unique axis *c* No. 13 (48)
(1) *hkl*: (2) $\overline{hk}l$: −010/2

$P2_1/c$ $P12_1/c1$ Unique axis *b* No. 14 (49)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −011/2

$P2_1/c$ $P12_1/n1$ Unique axis *b* No. 14 (50)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −111/2

$P2_1/c$ $P12_1/a1$ Unique axis *b* No. 14 (51)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −110/2

$P2_1/c$ $P112_1/a$ Unique axis *c* No. 14 (52)
(1) *hkl*: (2) $\overline{hk}l$: −101/2

$P2_1/c$ $P112_1/n$ Unique axis *c* No. 14 (53)
(1) *hkl*: (2) $\overline{hk}l$: −111/2

$P2_1/c$ $P112_1/b$ Unique axis *c* No. 14 (54)
(1) *hkl*: (2) $\overline{hk}l$: −011/2

$C2/c$ $C12/c1$ Unique axis *b* No. 15 (55)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −001/2

$C2/c$ $A12/n1$ Unique axis *b* No. 15 (56)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −101/2

$C2/c$ $I12/a1$ Unique axis *b* No. 15 (57)
(1) *hkl*: (2) $\bar{h}k\bar{l}$: −100/2

$C2/c$ $A112/a$ Unique axis *c* No. 15 (58)
(1) *hkl*: (2) $\overline{hk}l$: −100/2

$C2/c$ $B112/n$ Unique axis *c* No. 15 (59)
(1) *hkl*: (2) $\overline{hk}l$: −110/2

$C2/c$ $I112/b$ Unique axis *c* No. 15 (60)
(1) *hkl*: (2) $\overline{hk}l$: −010/2

Point group: 222 Orthorhombic Laue group: *mmm*

$P222$ No. 16 (61)
(1) *hkl*: (2) $\bar{h}kl$: (3) $\bar{h}t\bar{l}$: (4) $h\bar{k}\bar{l}$:

$P222_1$ No. 17 (62)
(1) *hkl*: (2) $\bar{h}kl$: −001/2 (3) $\bar{h}k\bar{l}$: −001/2 (4) $h\bar{k}\bar{l}$:

$P2_12_12$ No. 18 (63)
(1) *hkl*: (2) $\bar{h}kl$: (3) $\bar{h}k\bar{l}$: −110/2 (4) $h\bar{k}\bar{l}$: −110/2

$P2_12_12_1$ No. 19 (64)
(1) *hkl*: (2) $\bar{h}kl$: −101/2 (3) $\bar{h}k\bar{l}$: −011/2 (4) $h\bar{k}\bar{l}$: −110/2

$C222_1$ No. 20 (65)
(1) *hkl*: (2) $\bar{h}kl$: −001/2 (3) $\bar{h}k\bar{l}$: −001/2 (4) $h\bar{k}\bar{l}$:

$C222$ No. 21 (66)
(1) *hkl*: (2) $\bar{h}kl$: (3) $\bar{h}k\bar{l}$: (4) $h\bar{k}\bar{l}$:

$F222$ No. 22 (67)
(1) *hkl*: (2) $\bar{h}kl$: (3) $\bar{h}k\bar{l}$: (4) $h\bar{k}\bar{l}$:

$I222$ No. 23 (68)
(1) *hkl*: (2) $\bar{h}kl$: (3) $\bar{h}t\bar{l}$: (4) $h\bar{k}\bar{l}$:

Table A1.4.4.1 (*cont.*)

I$2_12_12_1$ No. 24 (69)
(1) *hkl*: (2) $\overline{hk}l$: −101/2 (3) $\overline{h}k\overline{l}$: −011/2 (4) $h\overline{k}l$: −110/2

Point group: *mm*2 Orthorhombic Laue group: *mmm*

*Pmm*2 No. 25 (70)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: (4) $\overline{h}kl$:

Pmc2_1 No. 26 (71)
(1) *hkl*: (2) $\overline{hk}l$: −001/2 (3) $h\overline{k}l$: −001/2 (4) $\overline{h}kl$:

*Pcc*2 No. 27 (72)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −001/2 (4) $\overline{h}kl$: −001/2

*Pma*2 No. 28 (73)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −100/2 (4) $\overline{h}kl$: −100/2

Pca2_1 No. 29 (74)
(1) *hkl*: (2) $\overline{hk}l$: −001/2 (3) $h\overline{k}l$: −100/2 (4) $\overline{h}kl$: −101/2

*Pnc*2 No. 30 (75)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −011/2 (4) $\overline{h}kl$: −011/2

Pmn2_1 No. 31 (76)
(1) *hkl*: (2) $\overline{hk}l$: −101/2 (3) $h\overline{k}l$: −101/2 (4) $\overline{h}kl$:

*Pba*2 No. 32 (77)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −110/2 (4) $\overline{h}kl$: −110/2

Pna2_1 No. 33 (78)
(1) *hkl*: (2) $\overline{hk}l$: −001/2 (3) $h\overline{k}l$: −110/2 (4) $\overline{h}kl$: −111/2

*Pnn*2 No. 34 (79)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −111/2 (4) $\overline{h}kl$: −111/2

*Cmm*2 No. 35 (80)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: (4) $\overline{h}kl$:

Cmc2_1 No. 36 (81)
(1) *hkl*: (2) $\overline{hk}l$: −001/2 (3) $h\overline{k}l$: −001/2 (4) $\overline{h}kl$:

*Ccc*2 No. 37 (82)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −001/2 (4) $\overline{h}kl$: −001/2

*Amm*2 No. 38 (83)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: (4) $\overline{h}kl$:

*Abm*2 No. 39 (84)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −010/2 (4) $\overline{h}kl$: −010/2

*Ama*2 No. 40 (85)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −100/2 (4) $\overline{h}kl$: −100/2

*Aba*2 No. 41 (86)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −110/2 (4) $\overline{h}kl$: −110/2

*Fmm*2 No. 42 (87)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: (4) $\overline{h}kl$:

*Fdd*2 No. 43 (88)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −313/4 (4) $\overline{h}kl$: −133/4

*Imm*2 No. 44 (89)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: (4) $\overline{h}kl$:

*Iba*2 No. 45 (90)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −110/2 (4) $\overline{h}kl$: −110/2

*Ima*2 No. 46 (91)
(1) *hkl*: (2) $\overline{hk}l$: (3) $h\overline{k}l$: −100/2 (4) $\overline{h}kl$: −100/2

Point group: *mmm* Orthorhombic Laue group: *mmm*

Pmmm No. 47 (92)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:

Pnnn Origin 1 No. 48 (93)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:
(5) $\overline{hk}l$: −111/2 (6) $hk\overline{l}$: −111/2 (7) $h\overline{k}l$: −111/2 (8) $\overline{h}kl$: −111/2

Pnnn Origin 2 No. 48 (94)
(1) *hkl*: (2) $\overline{hk}l$: −110/2 (3) $\overline{h}k\overline{l}$: −101/2 (4) $h\overline{k}l$: −011/2

Pccm No. 49 (95)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: −001/2 (4) $h\overline{k}l$: −001/2

Pban Origin 1 No. 50 (96)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:
(5) $\overline{hk}l$: −110/2 (6) $hk\overline{l}$: −110/2 (7) $h\overline{k}l$: −110/2 (8) $\overline{h}kl$: −110/2

Pban Origin 2 No. 50 (97)
(1) *hkl*: (2) $\overline{hk}l$: −110/2 (3) $\overline{h}k\overline{l}$: −100/2 (4) $h\overline{k}l$: −010/2

Pmma No. 51 (98)
(1) *hkl*: (2) $\overline{hk}l$: −100/2 (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$: −100/2

Pnna No. 52 (99)
(1) *hkl*: (2) $\overline{hk}l$: −100/2 (3) $\overline{h}k\overline{l}$: −111/2 (4) $h\overline{k}l$: −011/2

Pmna No. 53 (100)
(1) *hkl*: (2) $\overline{hk}l$: −101/2 (3) $\overline{h}k\overline{l}$: −101/2 (4) $h\overline{k}l$:

Pcca No. 54 (101)
(1) *hkl*: (2) $\overline{hk}l$: −100/2 (3) $\overline{h}k\overline{l}$: −001/2 (4) $h\overline{k}l$: −101/2

Pbam No. 55 (102)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: −110/2 (4) $h\overline{k}l$: −110/2

Pccn No. 56 (103)
(1) *hkl*: (2) $\overline{hk}l$: −110/2 (3) $\overline{h}k\overline{l}$: −011/2 (4) $h\overline{k}l$: −101/2

Pbcm No. 57 (104)
(1) *hkl*: (2) $\overline{hk}l$: −001/2 (3) $\overline{h}k\overline{l}$: −011/2 (4) $h\overline{k}l$: −010/2

Pnnm No. 58 (105)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: −111/2 (4) $h\overline{k}l$: −111/2

Pmmn Origin 1 No. 59 (106)
(1) *hkl*: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: −110/2 (4) $h\overline{k}l$: −110/2
(5) $\overline{hk}l$: −110/2 (6) $hk\overline{l}$: −110/2 (7) $h\overline{k}l$: (8) $\overline{h}kl$:

Pmmn Origin 2 No. 59 (107)
(1) *hkl*: (2) $\overline{hk}l$: −110/2 (3) $\overline{h}k\overline{l}$: −010/2 (4) $h\overline{k}l$: −100/2

Table A1.4.4.1 (*cont.*)

Pbcn No. 60 (108)
(1) hkl: (2) $\overline{hk}l$: −111/2 (3) $\overline{h}k\overline{l}$: −001/2 (4) $h\overline{k}l$: −110/2

Pbca No. 61 (109)
(1) hkl: (2) $\overline{hk}l$: −101/2 (3) $\overline{h}k\overline{l}$: −011/2 (4) $h\overline{k}l$: −110/2

Pnma No. 62 (110)
(1) hkl: (2) $\overline{hk}l$: −101/2 (3) $\overline{h}k\overline{l}$: −010/2 (4) $h\overline{k}l$: −111/2

Cmcm No. 63 (111)
(1) hkl: (2) $\overline{hk}l$: −001/2 (3) $\overline{h}k\overline{l}$: −001/2 (4) $h\overline{k}l$:

Cmca No. 64 (112)
(1) hkl: (2) $\overline{hk}l$: −011/2 (3) $\overline{h}k\overline{l}$: −011/2 (4) $h\overline{k}l$:

Cmmm No. 65 (113)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:

Cccm No. 66 (114)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: −001/2 (4) $h\overline{k}l$: −001/2

Cmma No. 67 (115)
(1) hkl: (2) $\overline{hk}l$: −010/2 (3) $\overline{h}k\overline{l}$: −010/2 (4) $h\overline{k}l$:

Ccca Origin 1 No. 68 (116)
(1) hkl: (2) $\overline{hk}l$: −110/2 (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$: −110/2
(5) $\overline{hk}l$: −011/2 (6) $hk\overline{l}$: −101/2 (7) $h\overline{k}l$: −011/2 (8) $\overline{h}kl$: −101/2

Ccca Origin 2 No. 68 (117)
(1) hkl: (2) $\overline{hk}l$: −100/2 (3) $\overline{h}k\overline{l}$: −001/2 (4) $h\overline{k}l$: −101/2

Fmmm No. 69 (118)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:

Fddd Origin 1 No. 70 (119)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:
(5) $\overline{hk}l$: −111/4 (6) $hk\overline{l}$: −111/4 (7) $h\overline{k}l$: −111/4 (8) $\overline{h}kl$: −111/4

Fddd Origin 2 No. 70 (120)
(1) hkl: (2) $\overline{hk}l$: −330/4 (3) $\overline{h}k\overline{l}$: −303/4 (4) $h\overline{k}l$: −033/4

Immm No. 71 (121)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: (4) $h\overline{k}l$:

Ibam No. 72 (122)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{h}k\overline{l}$: −110/2 (4) $h\overline{k}l$: −110/2

Ibca No. 73 (123)
(1) hkl: (2) $\overline{hk}l$: −101/2 (3) $\overline{h}k\overline{l}$: −011/2 (4) $h\overline{k}l$: −110/2

Imma No. 74 (124)
(1) hkl: (2) $\overline{hk}l$: −010/2 (3) $\overline{h}k\overline{l}$: −010/2 (4) $h\overline{k}l$:

Point group: 4 Tetragonal Laue group: 4/*m*

*P*4 No. 75 (125)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:

$P4_1$ No. 76 (126)
(1) hkl: (2) $\overline{hk}l$: −001/2 (3) $k\overline{h}l$: −001/4 (4) $\overline{k}hl$: −003/4

$P4_2$ No. 77 (127)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: −001/2 (4) $\overline{k}hl$: −001/2

$P4_3$ No. 78 (128)
(1) hkl: (2) $\overline{hk}l$: −001/2 (3) $k\overline{h}l$: −003/4 (4) $\overline{k}hl$: −001/4

*I*4 No. 79 (129)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:

$I4_1$ No. 80 (130)
(1) hkl: (2) $\overline{hk}l$: −111/2 (3) $k\overline{h}l$: −021/4 (4) $\overline{k}hl$: −203/4

Point group: $\bar{4}$ Tetragonal Laue group: 4/*m*

$P\bar{4}$ No. 81 (131)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{k}h\overline{l}$: (4) $k\overline{h}\overline{l}$:

$I\bar{4}$ No. 82 (132)
(1) hkl: (2) $\overline{hk}l$: (3) $\overline{k}h\overline{l}$: (4) $k\overline{h}\overline{l}$:

Point group: 4/*m* Tetragonal Laue group: 4/*m*

$P4/m$ No. 83 (133)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:

$P4_2/m$ No. 84 (134)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: −001/2 (4) $\overline{k}hl$: −001/2

$P4/n$ Origin 1 No. 85 (135)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: −110/2 (4) $\overline{k}hl$: −110/2
(5) $\overline{hk}l$: −110/2 (6) $hk\overline{l}$: −110/2 (7) $\overline{k}h\overline{l}$: (8) $k\overline{h}\overline{l}$:

$P4/n$ Origin 2 No. 85 (136)
(1) hkl: (2) $\overline{hk}l$: −110/2 (3) $k\overline{h}l$: −100/2 (4) $\overline{k}hl$: −010/2

$P4_2/n$ Origin 1 No. 86 (137)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: −111/2 (4) $\overline{k}hl$: −111/2
(5) $\overline{hk}l$: −111/2 (6) $hk\overline{l}$: −111/2 (7) $\overline{k}h\overline{l}$: (8) $k\overline{h}\overline{l}$:

$P4_2/n$ Origin 2 No. 86 (138)
(1) hkl: (2) $\overline{hk}l$: −110/2 (3) $k\overline{h}l$: −011/2 (4) $\overline{k}hl$: −101/2

$I4/m$ No. 87 (139)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:

$I4_1/a$ Origin 1 No. 88 (140)
(1) hkl: (2) $\overline{hk}l$: −111/2 (3) $k\overline{h}l$: −021/4 (4) $\overline{k}hl$: −203/4
(5) $\overline{hk}l$: −021/4 (6) $hk\overline{l}$: −203/4 (7) $\overline{k}h\overline{l}$: (8) $k\overline{h}\overline{l}$: −111/2

$I4_1/a$ Origin 2 No. 88 (141)
(1) hkl: (2) $\overline{hk}l$: −101/2 (3) $k\overline{h}l$: −311/4 (4) $\overline{k}hl$: −333/4

Point group: 422 Tetragonal Laue group: 4/*mmm*

*P*422 No. 89 (142)
(1) hkl: (2) $\overline{hk}l$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$: (6) $h\overline{k}\overline{l}$: (7) $kh\overline{l}$: (8) $\overline{kh}\overline{l}$:

Table A1.4.4.1 (*cont.*)

$P42_12$ No. 90 (143)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −110/2	(4) $\overline{k}hl$: −110/2
(5) $\overline{h}k\overline{l}$: −110/2	(6) $h\overline{k}\overline{l}$: −110/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$P4_122$ No. 91 (144)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −001/2	(3) $k\overline{h}l$: −001/4	(4) $\overline{k}hl$: −003/4
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$: −001/2	(7) $kh\overline{l}$: −003/4	(8) $\overline{k}\overline{h}l$: −001/4

$P4_12_12$ No. 92 (145)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −001/2	(3) $k\overline{h}l$: −221/4	(4) $\overline{k}hl$: −223/4
(5) $\overline{h}k\overline{l}$: −221/4	(6) $h\overline{k}\overline{l}$: −223/4	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$: −001/2

$P4_222$ No. 93 (146)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2

$P4_22_12$ No. 94 (147)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $\overline{h}k\overline{l}$: −111/2	(6) $h\overline{k}\overline{l}$: −111/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$P4_322$ No. 95 (148)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −001/2	(3) $k\overline{h}l$: −003/4	(4) $\overline{k}hl$: −001/4
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$: −001/2	(7) $kh\overline{l}$: −001/4	(8) $\overline{k}\overline{h}l$: −003/4

$P4_32_12$ No. 96 (149)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −001/2	(3) $k\overline{h}l$: −223/4	(4) $\overline{k}hl$: −221/4
(5) $\overline{h}k\overline{l}$: −223/4	(6) $h\overline{k}\overline{l}$: −221/4	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$: −001/2

$I422$ No. 97 (150)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$I4_122$ No. 98 (151)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −111/2	(3) $k\overline{h}l$: −021/4	(4) $\overline{k}hl$: −203/4
(5) $\overline{h}k\overline{l}$: −203/4	(6) $h\overline{k}\overline{l}$: −021/4	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$:

Point group: $4mm$ Tetragonal Laue group: $4/mmm$

$P4mm$ No. 99 (152)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $h\overline{k}l$:	(6) $\overline{h}kl$:	(7) $\overline{k}\overline{h}l$:	(8) khl:

$P4bm$ No. 100 (153)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $h\overline{k}l$: −110/2	(6) $\overline{h}kl$: −110/2	(7) $\overline{k}\overline{h}l$: −110/2	(8) khl: −110/2

$P4_2cm$ No. 101 (154)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $h\overline{k}l$: −001/2	(6) $\overline{h}kl$: −001/2	(7) $\overline{k}\overline{h}l$:	(8) khl:

$P4_2nm$ No. 102 (155)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $h\overline{k}l$: −111/2	(6) $\overline{h}kl$: −111/2	(7) $\overline{k}\overline{h}l$:	(8) khl:

$P4cc$ No. 103 (156)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $h\overline{k}l$: −001/2	(6) $\overline{h}kl$: −001/2	(7) $\overline{k}\overline{h}l$: −001/2	(8) khl: −001/2

$P4nc$ No. 104 (157)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $h\overline{k}l$: −111/2	(6) $\overline{h}kl$: −111/2	(7) $\overline{k}\overline{h}l$: −111/2	(8) khl: −111/2

$P4_2mc$ No. 105 (158)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $h\overline{k}l$:	(6) $\overline{h}kl$:	(7) $\overline{k}\overline{h}l$: −001/2	(8) khl: −001/2

$P4_2bc$ No. 106 (159)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $h\overline{k}l$: −110/2	(6) $\overline{h}kl$: −110/2	(7) $\overline{k}\overline{h}l$: −111/2	(8) khl: −111/2

$I4mm$ No. 107 (160)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $h\overline{k}l$:	(6) $\overline{h}kl$:	(7) $\overline{k}\overline{h}l$:	(8) khl:

$I4cm$ No. 108 (161)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}l$:	(4) $\overline{k}hl$:
(5) $h\overline{k}l$: −001/2	(6) $\overline{h}kl$: −001/2	(7) $\overline{k}\overline{h}l$: −001/2	(8) khl: −001/2

$I4_1md$ No. 109 (162)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −111/2	(3) $k\overline{h}l$: −021/4	(4) $\overline{k}hl$: −203/4
(5) $h\overline{k}l$:	(6) $\overline{h}kl$: −111/2	(7) $\overline{k}\overline{h}l$: −203/4	(8) khl: −021/4

$I4_1cd$ No. 110 (163)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −111/2	(3) $k\overline{h}l$: −021/4	(4) $\overline{k}hl$: −203/4
(5) $h\overline{k}l$: −001/2	(6) $\overline{h}kl$: −110/2	(7) $\overline{k}\overline{h}l$: −201/4	(8) khl: −023/4

Point group: $\bar{4}2m$ Tetragonal Laue group: $4/mmm$

$P\bar{4}2m$ No. 111 (164)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $\overline{k}\overline{h}l$:	(8) khl:

$P\bar{4}2c$ No. 112 (165)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $\overline{h}k\overline{l}$: −001/2	(6) $h\overline{k}\overline{l}$: −001/2	(7) $\overline{k}\overline{h}l$: −001/2	(8) khl: −001/2

$P\bar{4}2_1m$ No. 113 (166)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $\overline{h}k\overline{l}$: −110/2	(6) $h\overline{k}\overline{l}$: −110/2	(7) $\overline{k}\overline{h}l$: −110/2	(8) khl: −110/2

$P\bar{4}2_1c$ No. 114 (167)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $\overline{h}k\overline{l}$: −111/2	(6) $h\overline{k}\overline{l}$: −111/2	(7) $\overline{k}\overline{h}l$: −111/2	(8) khl: −111/2

$P\bar{4}m2$ No. 115 (168)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $h\overline{k}l$:	(6) $\overline{h}kl$:	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$P\bar{4}c2$ No. 116 (169)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $h\overline{k}l$: −001/2	(6) $\overline{h}kl$: −001/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2

$P\bar{4}b2$ No. 117 (170)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $h\overline{k}l$: −110/2	(6) $\overline{h}kl$: −110/2	(7) $kh\overline{l}$: −110/2	(8) $\overline{k}\overline{h}l$: −110/2

$P\bar{4}n2$ No. 118 (171)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $h\overline{k}l$: −111/2	(6) $\overline{h}kl$: −111/2	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$: −111/2

$I\bar{4}m2$ No. 119 (172)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $h\overline{k}l$:	(6) $\overline{h}kl$:	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

Table A1.4.4.1 (*cont.*)

$I\bar{4}c2$ No. 120 (173)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $h\overline{k}\overline{l}$: −001/2	(6) $\overline{h}kl$: −001/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2

$I\bar{4}2m$ No. 121 (174)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $\overline{k}hl$:	(8) khl:

$I\bar{4}2d$ No. 122 (175)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $\overline{k}h\overline{l}$:	(4) $k\overline{h}\overline{l}$:
(5) $\overline{h}k\overline{l}$: −203/4	(6) $h\overline{k}\overline{l}$: −203/4	(7) $\overline{k}hl$: −021/4	(8) khl: −021/4

Point group: 4/*mmm* Tetragonal Laue group: 4/*mmm*

$P4/mmm$ No. 123 (176)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$P4/mcc$ No. 124 (177)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$: −001/2	(6) $h\overline{k}\overline{l}$: −001/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2

$P4/nbm$ Origin 1 No. 125 (178)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:
(9) $\overline{h}\overline{k}l$: −110/2	(10) $hk\overline{l}$: −110/2	(11) $\overline{k}h\overline{l}$: −110/2	(12) $k\overline{h}\overline{l}$: −110/2
(13) $h\overline{k}\overline{l}$: −110/2	(14) $\overline{h}kl$: −110/2	(15) $\overline{k}\overline{h}l$: −110/2	(16) khl: −110/2

$P4/nbm$ Origin 2 No. 125 (179)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −100/2	(4) $\overline{k}hl$: −010/2
(5) $\overline{h}k\overline{l}$: −100/2	(6) $h\overline{k}\overline{l}$: −010/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$: −110/2

$P4/nnc$ Origin 1 No. 126 (180)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:
(9) $\overline{h}\overline{k}l$: −111/2	(10) $hk\overline{l}$: −111/2	(11) $\overline{k}h\overline{l}$: −111/2	(12) $k\overline{h}\overline{l}$: −111/2
(13) $h\overline{k}\overline{l}$: −111/2	(14) $\overline{h}kl$: −111/2	(15) $\overline{k}\overline{h}l$: −111/2	(16) khl: −111/2

$P4/nnc$ Origin 2 No. 126 (181)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −100/2	(4) $\overline{k}hl$: −010/2
(5) $\overline{h}k\overline{l}$: −101/2	(6) $h\overline{k}\overline{l}$: −011/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −111/2

$P4/mbm$ No. 127 (182)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$: −110/2	(6) $h\overline{k}\overline{l}$: −110/2	(7) $kh\overline{l}$: −110/2	(8) $\overline{k}\overline{h}l$: −110/2

$P4/mnc$ No. 128 (183)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$:	(4) $\overline{k}hl$:
(5) $\overline{h}k\overline{l}$: −111/2	(6) $h\overline{k}\overline{l}$: −111/2	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$: −111/2

$P4/nmm$ Origin 1 No. 129 (184)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −110/2	(4) $\overline{k}hl$: −110/2
(5) $\overline{h}k\overline{l}$: −110/2	(6) $h\overline{k}\overline{l}$: −110/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:
(9) $\overline{h}\overline{k}l$: −110/2	(10) $hk\overline{l}$: –110/2	(11) $\overline{k}h\overline{l}$:	(12) $k\overline{h}\overline{l}$:
(13) $h\overline{k}\overline{l}$:	(14) $\overline{h}kl$:	(15) $\overline{k}\overline{h}l$: −110/2	(16) khl: −110/2

$P4/nmm$ Origin 2 No. 129 (185)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −100/2	(4) $\overline{k}hl$: −010/2
(5) $\overline{h}k\overline{l}$: −010/2	(6) $h\overline{k}\overline{l}$: −100/2	(7) $kh\overline{l}$: −110/2	(8) $\overline{k}\overline{h}l$:

$P4/ncc$ Origin 1 No. 130 (186)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −110/2	(4) $\overline{k}hl$: −110/2
(5) $\overline{h}k\overline{l}$: −111/2	(6) $h\overline{k}\overline{l}$: −111/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2
(9) $\overline{h}\overline{k}l$: −110/2	(10) $hk\overline{l}$: −110/2	(11) $\overline{k}h\overline{l}$:	(12) $k\overline{h}\overline{l}$:
(13) $h\overline{k}\overline{l}$: −001/2	(14) $\overline{h}kl$: −001/2	(15) $\overline{k}\overline{h}l$: −111/2	(16) khl: −111/2

$P4/ncc$ Origin 2 No. 130 (187)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −100/2	(4) $\overline{k}hl$: −010/2
(5) $\overline{h}k\overline{l}$: −011/2	(6) $h\overline{k}\overline{l}$: −101/2	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$: −001/2

$P4_2/mmc$ No. 131 (188)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2

$P4_2/mcm$ No. 132 (189)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $\overline{h}k\overline{l}$: −001/2	(6) $h\overline{k}\overline{l}$: −001/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$P4_2/nbc$ Origin 1 No. 133 (190)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $\overline{h}k\overline{l}$: −001/2	(6) $h\overline{k}\overline{l}$: −001/2	(7) $kh\overline{l}$: −110/2	(8) $\overline{k}\overline{h}l$: −110/2
(9) $\overline{h}\overline{k}l$: −111/2	(10) $hk\overline{l}$: −111/2	(11) $\overline{k}h\overline{l}$:	(12) $k\overline{h}\overline{l}$:
(13) $h\overline{k}\overline{l}$: −110/2	(14) $\overline{h}kl$: −110/2	(15) $\overline{k}\overline{h}l$: −001/2	(16) khl: −001/2

$P4_2/nbc$ Origin 2 No. 133 (191)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −101/2	(4) $\overline{k}hl$: −011/2
(5) $\overline{h}k\overline{l}$: −100/2	(6) $h\overline{k}\overline{l}$: −010/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −111/2

$P4_2/nnm$ Origin 1 No. 134 (192)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $\overline{h}k\overline{l}$:	(6) $h\overline{k}\overline{l}$:	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$: −111/2
(9) $\overline{h}\overline{k}l$: −111/2	(10) $hk\overline{l}$: −111/2	(11) $\overline{k}h\overline{l}$:	(12) $k\overline{h}\overline{l}$:
(13) $h\overline{k}\overline{l}$: −111/2	(14) $\overline{h}kl$: −111/2	(15) $\overline{k}\overline{h}l$:	(16) khl:

$P4_2/nnm$ Origin 2 No. 134 (193)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −101/2	(4) $\overline{k}hl$: −011/2
(5) $\overline{h}k\overline{l}$: −101/2	(6) $h\overline{k}\overline{l}$: −011/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$: −110/2

$P4_2/mbc$ No. 135 (194)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −001/2	(4) $\overline{k}hl$: −001/2
(5) $\overline{h}k\overline{l}$: −110/2	(6) $h\overline{k}\overline{l}$: −110/2	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$: −111/2

$P4_2/mnm$ No. 136 (195)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $\overline{h}k\overline{l}$: −111/2	(6) $h\overline{k}\overline{l}$: −111/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:

$P4_2/nmc$ Origin 1 No. 137 (196)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $\overline{h}k\overline{l}$: −111/2	(6) $h\overline{k}\overline{l}$: −111/2	(7) $kh\overline{l}$:	(8) $\overline{k}\overline{h}l$:
(9) $\overline{h}\overline{k}l$: −111/2	(10) $hk\overline{l}$: −111/2	(11) $\overline{k}h\overline{l}$:	(12) $k\overline{h}\overline{l}$:
(13) $h\overline{k}\overline{l}$:	(14) $\overline{h}kl$:	(15) $\overline{k}\overline{h}l$: −111/2	(16) khl: −111/2

$P4_2/nmc$ Origin 2 No. 137 (197)

(1) hkl:	(2) $\overline{h}\overline{k}l$: −110/2	(3) $k\overline{h}\overline{l}$: −101/2	(4) $\overline{k}hl$: −011/2
(5) $\overline{h}k\overline{l}$: −010/2	(6) $h\overline{k}\overline{l}$: −100/2	(7) $kh\overline{l}$: −111/2	(8) $\overline{k}\overline{h}l$: −001/2

$P4_2/ncm$ Origin 1 No. 138 (198)

(1) hkl:	(2) $\overline{h}\overline{k}l$:	(3) $k\overline{h}\overline{l}$: −111/2	(4) $\overline{k}hl$: −111/2
(5) $\overline{h}k\overline{l}$: −110/2	(6) $h\overline{k}\overline{l}$: −110/2	(7) $kh\overline{l}$: −001/2	(8) $\overline{k}\overline{h}l$: −001/2
(9) $\overline{h}\overline{k}l$: −111/2	(10) $hk\overline{l}$: –111/2	(11) $\overline{k}h\overline{l}$:	(12) $k\overline{h}\overline{l}$:
(13) $h\overline{k}\overline{l}$: −001/2	(14) $\overline{h}kl$: −001/2	(15) $\overline{k}\overline{h}l$: −110/2	(16) khl: −110/2

Table A1.4.4.1 (*cont.*)

$P4_2/ncm$ Origin 2 No. 138 (199)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: −110/2 (3) $k\overline{h}l$: −101/2 (4) $\overline{k}hl$: −011/2
(5) $\overline{h}k\overline{l}$: −011/2 (6) $h\overline{k}\overline{l}$: −101/2 (7) $kh\overline{l}$: −110/2 (8) $\overline{k}\overline{h}l$:

$I4/mmm$ No. 139 (200)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:
(5) $\overline{h}kl$: (6) $h\overline{k}l$: (7) $kh\overline{l}$: (8) $\overline{k}\overline{h}l$:

$I4/mcm$ No. 140 (201)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: (3) $k\overline{h}l$: (4) $\overline{k}hl$:
(5) $\overline{h}kl$: −001/2 (6) $h\overline{k}l$: −001/2 (7) $kh\overline{l}$: −001/2 (8) $\overline{k}\overline{h}l$: −001/2

$I4_1/amd$ Origin 1 No. 141 (202)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: −111/2 (3) $k\overline{h}l$: −021/4 (4) $\overline{k}hl$: −203/4
(5) $\overline{h}kl$: −203/4 (6) $h\overline{k}l$: −021/4 (7) $kh\overline{l}$: −111/2 (8) $\overline{k}\overline{h}l$:
(9) $\overline{h}k\overline{l}$: −021/4 (10) $hk\overline{l}$: −203/4 (11) $\overline{k}h\overline{l}$: (12) $k\overline{h}l$: −111/2
(13) $h\overline{k}l$: −111/2 (14) $\overline{h}kl$: (15) $\overline{k}\overline{h}l$: −203/4 (16) khl: −021/4

$I4_1/amd$ Origin 2 No. 141 (203)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: −101/2 (3) $k\overline{h}l$: −131/4 (4) $\overline{k}hl$: −113/4
(5) $\overline{h}kl$: −101/2 (6) $h\overline{k}l$: (7) $kh\overline{l}$: −131/4 (8) $\overline{k}\overline{h}l$: −113/4

$I4_1/acd$ Origin 1 No. 142 (204)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: −111/2 (3) $k\overline{h}l$: −021/4 (4) $\overline{k}hl$: −203/4
(5) $\overline{h}kl$: −201/4 (6) $h\overline{k}l$: −023/4 (7) $kh\overline{l}$: −110/2 (8) $\overline{k}\overline{h}l$: −001/2
(9) $\overline{h}k\overline{l}$: −021/4 (10) $hk\overline{l}$: –203/4 (11) $\overline{k}h\overline{l}$: (12) $k\overline{h}l$: −111/2
(13) $h\overline{k}l$: −110/2 (14) $\overline{h}kl$: −001/2 (15) $\overline{k}\overline{h}l$: −201/4 (16) khl: −023/4

$I4_1/acd$ Origin 2 No. 142 (205)

(1) *hkl*: (2) $\overline{h}k\overline{l}$: −101/2 (3) $k\overline{h}l$: −131/4 (4) $\overline{k}hl$: −113/4
(5) $\overline{h}kl$: −100/2 (6) $h\overline{k}l$: −001/2 (7) $kh\overline{l}$: −133/4 (8) $\overline{k}\overline{h}l$: −111/4

Point group: 3 Trigonal Laue group: $\overline{3}$

$P3$ No. 143 (206)

(1) *hkl*: (2) *kil*: (3) *ihl*:

$P3_1$ No. 144 (207)

(1) *hkl*: (2) *kil*: −001/3 (3) *ihl*: −002/3

$P3_2$ No. 145 (208)

(1) *hkl*: (2) *kil*: −002/3 (3) *ihl*: −001/3

$R3$ (hexagonal axes) No. 146 (209)

(1) *hkl*: (2) *kil*: (3) *ihl*:

$R3$ (rhombohedral axes) No. 146 (210)

(1) *hkl*: (2) *klh*: (3) *lhk*:

Point group: $\overline{3}$ Trigonal Laue group: $\overline{3}$

$P\overline{3}$ No. 147 (211)

(1) *hkl*: (2) *kil*: (3) *ihl*:

$R\overline{3}$ (hexagonal axes) No. 148 (212)

(1) *hkl*: (2) *kil*: (3) *ihl*:

$R\overline{3}$ (rhombohedral axes) No. 148 (213)

(1) *hkl*: (2) *klh*: (3) *lhk*:

Point group: 32 Trigonal Laue group: $\overline{3}m$

$P312$ No. 149 (214)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) $\overline{k}h\overline{l}$: (5) $\overline{h}i\overline{l}$: (6) $\overline{i}k\overline{l}$:

$P321$ No. 150 (215)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) $kh\overline{l}$: (5) $hi\overline{l}$: (6) $ik\overline{l}$:

$P3_112$ No. 151 (216)

(1) *hkl*: (2) *kil*: −001/3 (3) *ihl*: −002/3
(4) $\overline{k}h\overline{l}$: −002/3 (5) $\overline{h}i\overline{l}$: −001/3 (6) $\overline{i}k\overline{l}$:

$P3_121$ No. 152 (217)

(1) *hkl*: (2) *kil*: −001/3 (3) *ihl*: −002/3
(4) $kh\overline{l}$: (5) $hi\overline{l}$: −002/3 (6) $ik\overline{l}$: −001/3

$P3_212$ No. 153 (218)

(1) *hkl*: (2) *kil*: −002/3 (3) *ihl*: −001/3
(4) $\overline{k}h\overline{l}$: −001/3 (5) $\overline{h}i\overline{l}$: −002/3 (6) $\overline{i}k\overline{l}$:

$P3_221$ No. 154 (219)

(1) *hkl*: (2) *kil*: −002/3 (3) *ihl*: −001/3
(4) $kh\overline{l}$: (5) $hi\overline{l}$: −001/3 (6) $ik\overline{l}$: −002/3

$R32$ (hexagonal axes) No. 155 (220)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) $kh\overline{l}$: (5) $hi\overline{l}$: (6) $ik\overline{l}$:

$R32$ (rhombohedral axes) No. 155 (221)

(1) *hkl*: (2) *klh*: (3) *lhk*:
(4) $\overline{k}h\overline{l}$: (5) $\overline{h}l\overline{k}$: (6) $\overline{l}k\overline{h}$:

Point group: 3*m* Trigonal Laue group: $\overline{3}m$

$P3m1$ No. 156 (222)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) $\overline{k}h\overline{l}$: (5) $\overline{h}i\overline{l}$: (6) $\overline{i}k\overline{l}$:

$P31m$ No. 157 (223)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) *khl*: (5) *hil*: (6) *ikl*:

$P3c1$ No. 158 (224)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) $\overline{k}h\overline{l}$: −001/2 (5) $\overline{h}i\overline{l}$: −001/2 (6) $\overline{i}k\overline{l}$: −001/2

$P31c$ No. 159 (225)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) *khl*: −001/2 (5) *hil*: −001/2 (6) *ikl*: −001/2

$R3m$ (hexagonal axes) No. 160 (226)

(1) *hkl*: (2) *kil*: (3) *ihl*:
(4) $\overline{k}h\overline{l}$: (5) $\overline{h}i\overline{l}$: (6) $\overline{i}k\overline{l}$:

$R3m$ (rhombohedral axes) No. 160 (227)

(1) *hkl*: (2) *klh*: (3) *lhk*:
(4) *khl*: (5) *hlk*: (6) *lkh*:

Table A1.4.4.1 (*cont.*)

*R*3*c* (hexagonal axes) No. 161 (228)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{k}h\overline{l}$: −001/2	(5) $\overline{h}i\overline{l}$: −001/2	(6) $\overline{i}k\overline{l}$: −001/2
*R*3*c* (rhombohedral axes) No. 161 (229)		
(1) *hkl*:	(2) *klh*:	(3) *lhk*:
(4) *khl*: −111/2	(5) *hlk*: −111/2	(6) *lkh*: −111/2

Point group: $\bar{3}m$ Trigonal Laue group: $\bar{3}m$

$P\bar{3}1m$ No. 162 (230)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{k}h\overline{l}$:	(5) $\overline{h}i\overline{l}$:	(6) $\overline{i}k\overline{l}$:
$P\bar{3}1c$ No. 163 (231)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{k}h\overline{l}$: −001/2	(5) $\overline{h}i\overline{l}$: −001/2	(6) $\overline{i}k\overline{l}$: −001/2
$P\bar{3}m1$ No. 164 (232)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $kh\bar{l}$:	(5) $hi\bar{l}$:	(6) $ik\bar{l}$:
$P\bar{3}c1$ No. 165 (233)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $kh\bar{l}$: −001/2	(5) $hi\bar{l}$: −001/2	(6) $ik\bar{l}$: −001/2
$R\bar{3}m$ (hexagonal axes) No. 166 (234)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $kh\bar{l}$:	(5) $hi\bar{l}$:	(6) $ik\bar{l}$:
$R\bar{3}m$ (rhombohedral axes) No. 166 (235)		
(1) *hkl*:	(2) *klh*:	(3) *lhk*:
(4) $\overline{k}h\overline{l}$:	(5) $\overline{h}l\overline{k}$:	(6) $\overline{l}k\overline{h}$:
$R\bar{3}c$ (hexagonal axes) No. 167 (236)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $kh\bar{l}$: −001/2	(5) $hi\bar{l}$: −001/2	(6) $ik\bar{l}$: −001/2
$R\bar{3}c$ (rhombohedral axes) No. 167 (237)		
(1) *hkl*:	(2) *klh*:	(3) *lhk*:
(4) $\overline{k}h\overline{l}$: −111/2	(5) $\overline{h}l\overline{k}$: −111/2	(6) $\overline{l}k\overline{h}$: −111/2

Point group: 6 Hexagonal Laue group: 6/*m*

*P*6 No. 168 (238)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$:	(6) $\overline{i}\overline{h}l$:
$P6_1$ No. 169 (239)		
(1) *hkl*:	(2) *kil*: −001/3	(3) *ihl*: −002/3
(4) $\overline{h}\overline{k}l$: −001/2	(5) $\overline{k}il$: −005/6	(6) $\overline{i}\overline{h}l$: −001/6
$P6_5$ No. 170 (240)		
(1) *hkl*:	(2) *kil*: −002/3	(3) *ihl*: −001/3
(4) $\overline{h}\overline{k}l$: −001/2	(5) $\overline{k}il$: −001/6	(6) $\overline{i}\overline{h}l$: −005/6
$P6_2$ No. 171 (241)		
(1) *hkl*:	(2) *kil*: −002/3	(3) *ihl*: −001/3
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$: −002/3	(6) $\overline{i}\overline{h}l$: −001/3
$P6_4$ No. 172 (242)		
(1) *hkl*:	(2) *kil*: −001/3	(3) *ihl*: −002/3
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$: −001/3	(6) $\overline{i}\overline{h}l$: −002/3
$P6_3$ No. 173 (243)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{h}\overline{k}l$: −001/2	(5) $\overline{k}il$: −001/2	(6) $\overline{i}\overline{h}l$: −001/2

Point group: $\bar{6}$ Hexagonal Laue group: 6/*m*

$P\bar{6}$ No. 174 (244)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $hk\bar{l}$:	(5) $ki\bar{l}$:	(6) $ih\bar{l}$:

Point group: 6/*m* Hexagonal Laue group: 6/*m*

$P6/m$ No. 175 (245)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$:	(6) $\overline{i}\overline{h}l$:
$P6_3/m$ No. 176 (246)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{h}\overline{k}l$: −001/2	(5) $\overline{k}il$: −001/2	(6) $\overline{i}\overline{h}l$: −001/2

Point group: 622 Hexagonal Laue group: 6/*mmm*

*P*622 No. 177 (247)		
(1) *hkl*:	(2) *kil*:	(3) *ihl*:
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$:	(6) $\overline{i}\overline{h}l$:
(7) $kh\bar{l}$:	(8) $hi\bar{l}$:	(9) $ik\bar{l}$:
(10) $\overline{k}h\overline{l}$:	(11) $\overline{h}i\overline{l}$:	(12) $\overline{i}k\overline{l}$:
$P6_122$ No. 178 (248)		
(1) *hkl*:	(2) *kil*: −001/3	(3) *ihl*: −002/3
(4) $\overline{h}\overline{k}l$: −001/2	(5) $\overline{k}il$: −005/6	(6) $\overline{i}\overline{h}l$: −001/6
(7) $kh\bar{l}$: −001/3	(8) $hi\bar{l}$:	(9) $ik\bar{l}$: −002/3
(10) $\overline{k}h\overline{l}$: −005/6	(11) $\overline{h}i\overline{l}$: −001/2	(12) $\overline{i}k\overline{l}$: −001/6
$P6_522$ No. 179 (249)		
(1) *hkl*:	(2) *kil*: −002/3	(3) *ihl*: −001/3
(4) $\overline{h}\overline{k}l$: −001/2	(5) $\overline{k}il$: −001/6	(6) $\overline{i}\overline{h}l$: −005/6
(7) $kh\bar{l}$: −002/3	(8) $hi\bar{l}$:	(9) $ik\bar{l}$: −001/3
(10) $\overline{k}h\overline{l}$: −001/6	(11) $\overline{h}i\overline{l}$: −001/2	(12) $\overline{i}k\overline{l}$: −005/6
$P6_222$ No. 180 (250)		
(1) *hkl*:	(2) *kil*: −002/3	(3) *ihl*: −001/3
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$: −002/3	(6) $\overline{i}\overline{h}l$: −001/3
(7) $kh\bar{l}$: −002/3	(8) $hi\bar{l}$:	(9) $ik\bar{l}$: −001/3
(10) $\overline{k}h\overline{l}$: −002/3	(11) $\overline{h}i\overline{l}$:	(12) $\overline{i}k\overline{l}$: −001/3
$P6_422$ No. 181 (251)		
(1) *hkl*:	(2) *kil*: −001/3	(3) *ihl*: −002/3
(4) $\overline{h}\overline{k}l$:	(5) $\overline{k}il$: −001/3	(6) $\overline{i}\overline{h}l$: −002/3
(7) $kh\bar{l}$: −001/3	(8) $hi\bar{l}$:	(9) $ik\bar{l}$: −002/3
(10) $\overline{k}h\overline{l}$: −001/3	(11) $\overline{h}i\overline{l}$:	(12) $\overline{i}k\overline{l}$: −002/3

Table A1.4.4.1 (*cont.*)

$P6_322$ No. 182 (252)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$: −001/2	(5) $\overline{ki}l$: −001/2	(6) $\overline{ih}l$: −001/2
(7) $kh\bar{l}$:	(8) $hi\bar{l}$:	(9) $ik\bar{l}$:
(10) $\overline{kh}l$: −001/2	(11) $\overline{hi}l$: −001/2	(12) $\overline{ik}l$: −001/2

Point group: 6*mm* Hexagonal Laue group: 6/*mmm*

$P6mm$ No. 183 (253)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$:	(5) $\overline{ki}l$:	(6) $\overline{ih}l$:
(7) $\overline{kh}l$:	(8) $\overline{hi}l$:	(9) $\overline{ik}l$:
(10) khl:	(11) hil:	(12) ikl:

$P6cc$ No. 184 (254)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$:	(5) $\overline{ki}l$:	(6) $\overline{ih}l$:
(7) $\overline{kh}l$: −001/2	(8) $\overline{hi}l$: −001/2	(9) $\overline{ik}l$: −001/2
(10) khl: −001/2	(11) hil: −001/2	(12) ikl: −001/2

$P6_3cm$ No. 185 (255)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$: −001/2	(5) $\overline{ki}l$: −001/2	(6) $\overline{ih}l$: −001/2
(7) $\overline{kh}l$: −001/2	(8) $\overline{hi}l$: −001/2	(9) $\overline{ik}l$: −001/2
(10) khl:	(11) hil:	(12) ikl:

$P6_3mc$ No. 186 (256)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$: −001/2	(5) $\overline{ki}l$: −001/2	(6) $\overline{ih}l$: −001/2
(7) $\overline{kh}l$:	(8) $\overline{hi}l$:	(9) $\overline{ik}l$:
(10) khl: −001/2	(11) hil: −001/2	(12) ikl: −001/2

Point group: $\bar{6}m2$ Hexagonal Laue group: 6/*mmm*

$P\bar{6}m2$ No. 187 (257)

(1) hkl:	(2) kil:	(3) ihl:
(4) $hk\bar{l}$:	(5) $ki\bar{l}$:	(6) $ih\bar{l}$:
(7) $\overline{kh}l$:	(8) $\overline{hi}l$:	(9) $\overline{ik}l$:
(10) $\overline{kh}\bar{l}$:	(11) $\overline{hi}\bar{l}$:	(12) $\overline{ik}\bar{l}$:

$P\bar{6}c2$ No. 188 (258)

(1) hkl:	(2) kil:	(3) ihl:
(4) $hk\bar{l}$: −001/2	(5) $ki\bar{l}$: −001/2	(6) $ih\bar{l}$: −001/2
(7) $\overline{kh}l$: −001/2	(8) $\overline{hi}l$: −001/2	(9) $\overline{ik}l$: −001/2
(10) $\overline{kh}\bar{l}$:	(11) $\overline{hi}\bar{l}$:	(12) $\overline{ik}\bar{l}$:

$P\bar{6}2m$ No. 189 (259)

(1) hkl:	(2) kil:	(3) ihl:
(4) $hk\bar{l}$:	(5) $ki\bar{l}$:	(6) $ih\bar{l}$:
(7) $kh\bar{l}$:	(8) $hi\bar{l}$:	(9) $ik\bar{l}$:
(10) khl:	(11) hil:	(12) ikl:

$P\bar{6}2c$ No. 190 (260)

(1) hkl:	(2) kil:	(3) ihl:
(4) $hk\bar{l}$: −001/2	(5) $ki\bar{l}$: −001/2	(6) $ih\bar{l}$: −001/2
(7) $kh\bar{l}$:	(8) $hi\bar{l}$:	(9) $ik\bar{l}$:
(10) khl: −001/2	(11) hil: −001/2	(12) ikl: −001/2

Point group: 6/*mmm* Hexagonal Laue group: 6/*mmm*

$P6/mmm$ No. 191 (261)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$:	(5) $\overline{ki}l$:	(6) $\overline{ih}l$:
(7) $kh\bar{l}$:	(8) $hi\bar{l}$:	(9) $ik\bar{l}$:
(10) $\overline{kh}l$:	(11) $\overline{hi}l$:	(12) $\overline{ik}l$:

$P6/mcc$ No. 192 (262)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$:	(5) $\overline{ki}l$:	(6) $\overline{ih}l$:
(7) $kh\bar{l}$: −001/2	(8) $hi\bar{l}$: −001/2	(9) $ik\bar{l}$: −001/2
(10) $\overline{kh}l$: −001/2	(11) $\overline{hi}l$: −001/2	(12) $\overline{ik}l$: −001/2

$P6_3/mcm$ No. 193 (263)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$: −001/2	(5) $\overline{ki}l$: −001/2	(6) $\overline{ih}l$: −001/2
(7) $kh\bar{l}$: −001/2	(8) $hi\bar{l}$: −001/2	(9) $ik\bar{l}$: −001/2
(10) $\overline{kh}l$:	(11) $\overline{hi}l$:	(12) $\overline{ik}l$:

$P6_3/mmc$ No. 194 (264)

(1) hkl:	(2) kil:	(3) ihl:
(4) $\overline{hk}l$: −001/2	(5) $\overline{ki}l$: −001/2	(6) $\overline{ih}l$: −001/2
(7) $kh\bar{l}$:	(8) $hi\bar{l}$:	(9) $ik\bar{l}$:
(10) $\overline{kh}\bar{l}$: −001/2	(11) $\overline{hi}\bar{l}$: −001/2	(12) $\overline{ik}\bar{l}$: −001/2

Point group: 23 Cubic Laue group: $m\bar{3}$

$P23$ No. 195 (265)

(1) hkl:	(2) $\overline{hk}l$:	(3) $\overline{h}k\bar{l}$:	(4) $h\overline{k}\bar{l}$:
(5) klh:	(6) $\overline{kl}h$:	(7) $\overline{k}l\bar{h}$:	(8) $k\overline{lh}$:
(9) lhk:	(10) $\overline{lh}k$:	(11) $\overline{l}h\bar{k}$:	(12) $l\overline{hk}$:

$F23$ No. 196 (266)

(1) hkl:	(2) $\overline{hk}l$:	(3) $\overline{h}k\bar{l}$:	(4) $h\overline{k}\bar{l}$:
(5) klh:	(6) $\overline{kl}h$:	(7) $\overline{k}l\bar{h}$:	(8) $k\overline{lh}$:
(9) lhk:	(10) $\overline{lh}k$:	(11) $\overline{l}h\bar{k}$:	(12) $l\overline{hk}$:

$I23$ No. 197 (267)

(1) hkl:	(2) $\overline{hk}l$:	(3) $\overline{h}k\bar{l}$:	(4) $h\overline{k}\bar{l}$:
(5) klh:	(6) $\overline{kl}h$:	(7) $\overline{k}l\bar{h}$:	(8) $k\overline{lh}$:
(9) lhk:	(10) $\overline{lh}k$:	(11) $\overline{l}h\bar{k}$:	(12) $l\overline{hk}$:

$P2_13$ No. 198 (268)

(1) hkl:	(2) $\overline{hk}l$: −101/2	(3) $\overline{h}k\bar{l}$: −011/2	(4) $h\overline{k}\bar{l}$: −110/2
(5) klh:	(6) $\overline{kl}h$: −110/2	(7) $\overline{k}l\bar{h}$: −101/2	(8) $k\overline{lh}$: −011/2
(9) lhk:	(10) $\overline{lh}k$: −011/2	(11) $\overline{l}h\bar{k}$: −110/2	(12) $l\overline{hk}$: −101/2

$I2_13$ No. 199 (269)

(1) hkl:	(2) $\overline{hk}l$: −101/2	(3) $\overline{h}k\bar{l}$: −011/2	(4) $h\overline{k}\bar{l}$: −110/2
(5) klh:	(6) $\overline{kl}h$: −110/2	(7) $\overline{k}l\bar{h}$: −101/2	(8) $k\overline{lh}$: −011/2
(9) lhk:	(10) $\overline{lh}k$: −011/2	(11) $\overline{l}h\bar{k}$: −110/2	(12) $l\overline{hk}$: −101/2

Point group: $m\bar{3}$ Cubic Laue group: $m\bar{3}$

$Pm\bar{3}$ No. 200 (270)

(1) hkl:	(2) $\overline{hk}l$:	(3) $\overline{h}k\bar{l}$:	(4) $h\overline{k}\bar{l}$:
(5) klh:	(6) $\overline{kl}h$:	(7) $\overline{k}l\bar{h}$:	(8) $k\overline{lh}$:
(9) lhk:	(10) $\overline{lh}k$:	(11) $\overline{l}h\bar{k}$:	(12) $l\overline{hk}$:

Table A1.4.4.1 (*cont.*)

$Pn\bar{3}$ Origin 1 No. 201 (271)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) $\overline{h}k\overline{l}$: −111/2	(14) $hk\overline{l}$: −111/2	(15) $h\overline{k}l$: −111/2	(16) $\overline{h}kl$: −111/2
(17) $\overline{k}l\overline{h}$: −111/2	(18) $kl\overline{h}$: −111/2	(19) $k\overline{l}h$: −111/2	(20) $\overline{k}lh$: −111/2
(21) $\overline{l}h\overline{k}$: −111/2	(22) $lh\overline{k}$: −111/2	(23) $l\overline{h}k$: −111/2	(24) $\overline{l}hk$: −111/2

$Pn\bar{3}$ Origin 2 No. 201 (272)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −110/2	(3) $\overline{h}\overline{k}l$: −101/2	(4) $h\overline{k}\overline{l}$: −011/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −011/2	(7) $\overline{k}\overline{l}h$: −110/2	(8) $k\overline{l}\overline{h}$: −101/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −101/2	(11) $\overline{l}\overline{h}k$: −011/2	(12) $l\overline{h}\overline{k}$: −110/2

$Fm\bar{3}$ No. 202 (273)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:

$Fd\bar{3}$ Origin 1 No. 203 (274)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) $\overline{h}k\overline{l}$: −111/4	(14) $hk\overline{l}$: −111/4	(15) $h\overline{k}l$: −111/4	(16) $\overline{h}kl$: −111/4
(17) $\overline{k}l\overline{h}$: −111/4	(18) $kl\overline{h}$: −111/4	(19) $k\overline{l}h$: −111/4	(20) $\overline{k}lh$: −111/4
(21) $\overline{l}h\overline{k}$: −111/4	(22) $lh\overline{k}$: −111/4	(23) $l\overline{h}k$: −111/4	(24) $\overline{l}hk$: −111/4

$Fd\bar{3}$ Origin 2 No. 203 (275)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −110/4	(3) $\overline{h}\overline{k}l$: −101/4	(4) $h\overline{k}\overline{l}$: −011/4
(5) klh:	(6) $\overline{k}l\overline{h}$: −011/4	(7) $\overline{k}\overline{l}h$: −110/4	(8) $k\overline{l}\overline{h}$: −101/4
(9) lhk:	(10) $\overline{l}h\overline{k}$: −101/4	(11) $\overline{l}\overline{h}k$: −011/4	(12) $l\overline{h}\overline{k}$: −110/4

$Im\bar{3}$ No. 204 (276)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:

$Pa\bar{3}$ No. 205 (277)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −101/2	(3) $\overline{h}\overline{k}l$: −011/2	(4) $h\overline{k}\overline{l}$: −110/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −110/2	(7) $\overline{k}\overline{l}h$: −101/2	(8) $k\overline{l}\overline{h}$: −011/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −011/2	(11) $\overline{l}\overline{h}k$: −110/2	(12) $l\overline{h}\overline{k}$: −101/2

$Ia\bar{3}$ No. 206 (278)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −101/2	(3) $\overline{h}\overline{k}l$: −011/2	(4) $h\overline{k}\overline{l}$: −110/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −110/2	(7) $\overline{k}\overline{l}h$: −101/2	(8) $k\overline{l}\overline{h}$: −011/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −011/2	(11) $\overline{l}\overline{h}k$: −110/2	(12) $l\overline{h}\overline{k}$: −101/2

Point group: 432 Cubic Laue group: $m\bar{3}m$

$P432$ No. 207 (279)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) $kh\overline{l}$:	(14) $\overline{k}h\overline{l}$:	(15) $\overline{k}hl$:	(16) $k\overline{h}l$:
(17) $h\overline{l}k$:	(18) $\overline{h}lk$:	(19) $\overline{h}l\overline{k}$:	(20) $hl\overline{k}$:
(21) $\overline{l}kh$:	(22) $l\overline{k}h$:	(23) $lk\overline{h}$:	(24) $\overline{l}k\overline{h}$:

$P4_232$ No. 208 (280)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) $kh\overline{l}$: −111/2	(14) $\overline{k}h\overline{l}$: −111/2	(15) $\overline{k}hl$: −111/2	(16) $k\overline{h}l$: −111/2
(17) $h\overline{l}k$: −111/2	(18) $\overline{h}lk$: −111/2	(19) $\overline{h}l\overline{k}$: −111/2	(20) $hl\overline{k}$: −111/2
(21) $\overline{l}kh$: −111/2	(22) $l\overline{k}h$: −111/2	(23) $lk\overline{h}$: −111/2	(24) $\overline{l}k\overline{h}$: −111/2

$F432$ No. 209 (281)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) $kh\overline{l}$:	(14) $\overline{k}h\overline{l}$:	(15) $\overline{k}hl$:	(16) $k\overline{h}l$:
(17) $h\overline{l}k$:	(18) $\overline{h}lk$:	(19) $\overline{h}l\overline{k}$:	(20) $hl\overline{k}$:
(21) $\overline{l}kh$:	(22) $l\overline{k}h$:	(23) $lk\overline{h}$:	(24) $\overline{l}k\overline{h}$:

$F4_132$ No. 210 (282)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −011/2	(3) $\overline{h}\overline{k}l$: −110/2	(4) $h\overline{k}\overline{l}$: −101/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −101/2	(7) $\overline{k}\overline{l}h$: −011/2	(8) $k\overline{l}\overline{h}$: −110/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −110/2	(11) $\overline{l}\overline{h}k$: −101/2	(12) $l\overline{h}\overline{k}$: −011/2
(13) $kh\overline{l}$: −313/4	(14) $\overline{k}h\overline{l}$: −111/4	(15) $\overline{k}hl$: −133/4	(16) $k\overline{h}l$: −331/4
(17) $h\overline{l}k$: −313/4	(18) $\overline{h}lk$: −331/4	(19) $\overline{h}l\overline{k}$: −111/4	(20) $hl\overline{k}$: −133/4
(21) $\overline{l}kh$: −313/4	(22) $l\overline{k}h$: −133/4	(23) $lk\overline{h}$: −331/4	(24) $\overline{l}k\overline{h}$: −111/4

$I432$ No. 211 (283)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) $kh\overline{l}$:	(14) $\overline{k}h\overline{l}$:	(15) $\overline{k}hl$:	(16) $k\overline{h}l$:
(17) $h\overline{l}k$:	(18) $\overline{h}lk$:	(19) $\overline{h}l\overline{k}$:	(20) $hl\overline{k}$:
(21) $\overline{l}kh$:	(22) $l\overline{k}h$:	(23) $lk\overline{h}$:	(24) $\overline{l}k\overline{h}$:

$P4_332$ No. 212 (284)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −101/2	(3) $\overline{h}\overline{k}l$: −011/2	(4) $h\overline{k}\overline{l}$: −110/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −110/2	(7) $\overline{k}\overline{l}h$: −101/2	(8) $k\overline{l}\overline{h}$: −011/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −011/2	(11) $\overline{l}\overline{h}k$: −110/2	(12) $l\overline{h}\overline{k}$: −101/2
(13) $kh\overline{l}$: −133/4	(14) $\overline{k}h\overline{l}$: −111/4	(15) $\overline{k}hl$: −331/4	(16) $k\overline{h}l$: −313/4
(17) $h\overline{l}k$: −133/4	(18) $\overline{h}lk$: −313/4	(19) $\overline{h}l\overline{k}$: −111/4	(20) $hl\overline{k}$: −331/4
(21) $\overline{l}kh$: −133/4	(22) $l\overline{k}h$: −331/4	(23) $lk\overline{h}$: −313/4	(24) $\overline{l}k\overline{h}$: −111/4

$P4_132$ No. 213 (285)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −101/2	(3) $\overline{h}\overline{k}l$: −011/2	(4) $h\overline{k}\overline{l}$: −110/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −110/2	(7) $\overline{k}\overline{l}h$: −101/2	(8) $k\overline{l}\overline{h}$: −011/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −011/2	(11) $\overline{l}\overline{h}k$: −110/2	(12) $l\overline{h}\overline{k}$: −101/2
(13) $kh\overline{l}$: −311/4	(14) $\overline{k}h\overline{l}$: −333/4	(15) $\overline{k}hl$: −113/4	(16) $k\overline{h}l$: −131/4
(17) $h\overline{l}k$: −311/4	(18) $\overline{h}lk$: −131/4	(19) $\overline{h}l\overline{k}$: −333/4	(20) $hl\overline{k}$: −113/4
(21) $\overline{l}kh$: −311/4	(22) $l\overline{k}h$: −113/4	(23) $lk\overline{h}$: −131/4	(24) $\overline{l}k\overline{h}$: −333/4

$I4_132$ No. 214 (286)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −101/2	(3) $\overline{h}\overline{k}l$: −011/2	(4) $h\overline{k}\overline{l}$: −110/2
(5) klh:	(6) $\overline{k}l\overline{h}$: −110/2	(7) $\overline{k}\overline{l}h$: −101/2	(8) $k\overline{l}\overline{h}$: −011/2
(9) lhk:	(10) $\overline{l}h\overline{k}$: −011/2	(11) $\overline{l}\overline{h}k$: −110/2	(12) $l\overline{h}\overline{k}$: −101/2
(13) $kh\overline{l}$: −311/4	(14) $\overline{k}h\overline{l}$: −333/4	(15) $\overline{k}hl$: −113/4	(16) $k\overline{h}l$: −131/4
(17) $h\overline{l}k$: −311/4	(18) $\overline{h}lk$: −131/4	(19) $\overline{h}l\overline{k}$: −333/4	(20) $hl\overline{k}$: −113/4
(21) $\overline{l}kh$: −311/4	(22) $l\overline{k}h$: −113/4	(23) $lk\overline{h}$: −131/4	(24) $\overline{l}k\overline{h}$: −333/4

Point group: $\bar{4}3m$ Cubic Laue group: $m\bar{3}m$

$P\bar{4}3m$ No. 215 (287)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}\overline{k}l$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}l\overline{h}$:	(7) $\overline{k}\overline{l}h$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}h\overline{k}$:	(11) $\overline{l}\overline{h}k$:	(12) $l\overline{h}\overline{k}$:
(13) khl:	(14) $\overline{k}h\overline{l}$:	(15) $\overline{k}\overline{h}l$:	(16) $k\overline{h}\overline{l}$:
(17) hlk:	(18) $\overline{h}l\overline{k}$:	(19) $\overline{h}\overline{l}k$:	(20) $h\overline{l}\overline{k}$:
(21) lkh:	(22) $\overline{l}kh$:	(23) $\overline{l}\overline{k}h$:	(24) $l\overline{k}\overline{h}$:

Table A1.4.4.1 (*cont.*)

$F\bar{4}3m$ No. 216 (288)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) khl:	(14) $\bar{k}hl$:	(15) $\bar{k}\bar{h}l$:	(16) $k\bar{h}\bar{l}$:
(17) hlk:	(18) $\bar{h}lk$:	(19) $\bar{h}\bar{l}k$:	(20) $h\bar{l}\bar{k}$:
(21) lkh:	(22) $\bar{l}kh$:	(23) $\bar{l}\bar{k}h$:	(24) $l\bar{k}\bar{h}$:

$I\bar{4}3m$ No. 217 (289)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) khl:	(14) $\bar{k}hl$:	(15) $\bar{k}\bar{h}l$:	(16) $k\bar{h}\bar{l}$:
(17) hlk:	(18) $\bar{h}lk$:	(19) $\bar{h}\bar{l}k$:	(20) $h\bar{l}\bar{k}$:
(21) lkh:	(22) $\bar{l}kh$:	(23) $\bar{l}\bar{k}h$:	(24) $l\bar{k}\bar{h}$:

$P\bar{4}3n$ No. 218 (290)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) khl: −111/2	(14) $\bar{k}hl$: −111/2	(15) $\bar{k}\bar{h}l$: −111/2	(16) $k\bar{h}\bar{l}$: −111/2
(17) hlk: −111/2	(18) $\bar{h}lk$: −111/2	(19) $\bar{h}\bar{l}k$: −111/2	(20) $h\bar{l}\bar{k}$: −111/2
(21) lkh: −111/2	(22) $\bar{l}kh$: −111/2	(23) $\bar{l}\bar{k}h$: −111/2	(24) $l\bar{k}\bar{h}$: −111/2

$F\bar{4}3c$ No. 219 (291)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) khl: −111/2	(14) $\bar{k}hl$: −111/2	(15) $\bar{k}\bar{h}l$: −111/2	(16) $k\bar{h}\bar{l}$: −111/2
(17) hlk: −111/2	(18) $\bar{h}lk$: −111/2	(19) $\bar{h}\bar{l}k$: −111/2	(20) $h\bar{l}\bar{k}$: −111/2
(21) lkh: −111/2	(22) $\bar{l}kh$: −111/2	(23) $\bar{l}\bar{k}h$: −111/2	(24) $l\bar{k}\bar{h}$: −111/2

$I\bar{4}3d$ No. 220 (292)

(1) hkl:	(2) $\bar{h}k\bar{l}$: −010/2	(3) $\bar{h}\bar{k}l$: −011/2	(4) $h\bar{k}\bar{l}$: −001/2
(5) klh:	(6) $\bar{k}l\bar{h}$: −001/2	(7) $\bar{k}\bar{l}h$: −101/2	(8) $k\bar{l}\bar{h}$: −100/2
(9) lhk:	(10) $\bar{l}hk$: −100/2	(11) $\bar{l}\bar{h}k$: −110/2	(12) $l\bar{h}\bar{k}$: −010/2
(13) khl: −111/4	(14) $\bar{k}hl$: −311/4	(15) $\bar{k}\bar{h}l$: −313/4	(16) $k\bar{h}\bar{l}$: −113/4
(17) hlk: −111/4	(18) $\bar{h}lk$: −113/4	(19) $\bar{h}\bar{l}k$: −133/4	(20) $h\bar{l}\bar{k}$: −131/4
(21) lkh: −111/4	(22) $\bar{l}kh$: −131/4	(23) $\bar{l}\bar{k}h$: −331/4	(24) $l\bar{k}\bar{h}$: −311/4

Point group: $m\bar{3}m$ Cubic Laue group: $m\bar{3}m$

$Pm\bar{3}m$ No. 221 (293)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) $kh\bar{l}$:	(14) $\bar{k}h\bar{l}$:	(15) $\bar{k}\bar{h}\bar{l}$:	(16) $k\bar{h}\bar{l}$:
(17) $hl\bar{k}$:	(18) $\bar{h}l\bar{k}$:	(19) $\bar{h}\bar{l}\bar{k}$:	(20) $h\bar{l}\bar{k}$:
(21) $l\bar{k}h$:	(22) $\bar{l}\bar{k}h$:	(23) $l\bar{k}\bar{h}$:	(24) $\bar{l}\bar{k}\bar{h}$:

$Pn\bar{3}n$ Origin 1 No. 222 (294)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) $kh\bar{l}$:	(14) $\bar{k}h\bar{l}$:	(15) $\bar{k}\bar{h}\bar{l}$:	(16) $k\bar{h}\bar{l}$:
(17) $hl\bar{k}$:	(18) $\bar{h}l\bar{k}$:	(19) $\bar{h}\bar{l}\bar{k}$:	(20) $h\bar{l}\bar{k}$:
(21) $l\bar{k}h$:	(22) $\bar{l}\bar{k}h$:	(23) $l\bar{k}\bar{h}$:	(24) $\bar{l}\bar{k}\bar{h}$:
(25) $\bar{h}\bar{k}\bar{l}$: −111/2	(26) $h\bar{k}\bar{l}$: −111/2	(27) $\bar{h}k\bar{l}$: −111/2	(28) $\bar{h}\bar{k}l$: −111/2
(29) $\bar{k}\bar{l}\bar{h}$: −111/2	(30) $k\bar{l}\bar{h}$: −111/2	(31) $\bar{k}l\bar{h}$: −111/2	(32) $\bar{k}\bar{l}h$: −111/2
(33) $\bar{l}\bar{h}\bar{k}$: −111/2	(34) $l\bar{h}\bar{k}$: −111/2	(35) $\bar{l}h\bar{k}$: −111/2	(36) $\bar{l}\bar{h}k$: −111/2
(37) $\bar{k}\bar{h}\bar{l}$: −111/2	(38) $kh\bar{l}$: −111/2	(39) $k\bar{h}\bar{l}$: −111/2	(40) $\bar{k}h\bar{l}$: −111/2
(41) $\bar{h}l\bar{k}$: −111/2	(42) $h\bar{l}\bar{k}$: −111/2	(43) $hl\bar{k}$: −111/2	(44) $\bar{h}l\bar{k}$: −111/2
(45) $l\bar{k}\bar{h}$: −111/2	(46) $\bar{l}k\bar{h}$: −111/2	(47) $\bar{l}\bar{k}h$: −111/2	(48) lkh: −111/2

$Pn\bar{3}n$ Origin 2 No. 222 (295)

(1) hkl:	(2) $\bar{h}k\bar{l}$: −110/2	(3) $\bar{h}\bar{k}l$: −101/2	(4) $h\bar{k}\bar{l}$: −011/2
(5) klh:	(6) $\bar{k}l\bar{h}$: −011/2	(7) $\bar{k}\bar{l}h$: −110/2	(8) $k\bar{l}\bar{h}$: −101/2
(9) lhk:	(10) $\bar{l}hk$: −101/2	(11) $\bar{l}\bar{h}k$: −011/2	(12) $l\bar{h}\bar{k}$: −110/2
(13) $kh\bar{l}$: −001/2	(14) $\bar{k}h\bar{l}$: −111/2	(15) $\bar{k}\bar{h}\bar{l}$: −010/2	(16) $k\bar{h}\bar{l}$: −100/2
(17) $hl\bar{k}$: −001/2	(18) $\bar{h}l\bar{k}$: −100/2	(19) $\bar{h}\bar{l}\bar{k}$: −111/2	(20) $h\bar{l}\bar{k}$: −010/2
(21) $\bar{l}kh$: −001/2	(22) $l\bar{k}h$: −010/2	(23) $lk\bar{h}$: −100/2	(24) $\bar{l}\bar{k}\bar{h}$: −111/2

$Pm\bar{3}n$ No. 223 (296)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) $kh\bar{l}$: −111/2	(14) $\bar{k}h\bar{l}$: −111/2	(15) $\bar{k}\bar{h}\bar{l}$: −111/2	(16) $k\bar{h}\bar{l}$: −111/2
(17) $hl\bar{k}$: −111/2	(18) $\bar{h}l\bar{k}$: −111/2	(19) $\bar{h}\bar{l}\bar{k}$: −111/2	(20) $h\bar{l}\bar{k}$: −111/2
(21) $\bar{l}kh$: −111/2	(22) $l\bar{k}h$: −111/2	(23) $lk\bar{h}$: −111/2	(24) $\bar{l}\bar{k}\bar{h}$: −111/2

$Pn\bar{3}m$ Origin 1 No. 224 (297)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) $kh\bar{l}$: −111/2	(14) $\bar{k}h\bar{l}$: −111/2	(15) $\bar{k}\bar{h}\bar{l}$: −111/2	(16) $k\bar{h}\bar{l}$: −111/2
(17) $hl\bar{k}$: −111/2	(18) $\bar{h}l\bar{k}$: −111/2	(19) $\bar{h}\bar{l}\bar{k}$: −111/2	(20) $h\bar{l}\bar{k}$: −111/2
(21) $\bar{l}kh$: −111/2	(22) $l\bar{k}h$: −111/2	(23) $lk\bar{h}$: −111/2	(24) $\bar{l}\bar{k}\bar{h}$: −111/2
(25) $\bar{h}\bar{k}\bar{l}$: −111/2	(26) $h\bar{k}\bar{l}$: −111/2	(27) $\bar{h}k\bar{l}$: −111/2	(28) $\bar{h}\bar{k}l$: −111/2
(29) $\bar{k}\bar{l}\bar{h}$: −111/2	(30) $k\bar{l}\bar{h}$: −111/2	(31) $\bar{k}l\bar{h}$: −111/2	(32) $\bar{k}\bar{l}h$: −111/2
(33) $\bar{l}\bar{h}\bar{k}$: −111/2	(34) $l\bar{h}\bar{k}$: −111/2	(35) $\bar{l}h\bar{k}$: −111/2	(36) $\bar{l}\bar{h}k$: −111/2
(37) $\bar{k}\bar{h}\bar{l}$:	(38) $kh\bar{l}$:	(39) $k\bar{h}\bar{l}$:	(40) $\bar{k}h\bar{l}$:
(41) $\bar{h}l\bar{k}$:	(42) $h\bar{l}\bar{k}$:	(43) $hl\bar{k}$:	(44) $\bar{h}l\bar{k}$:
(45) $l\bar{k}\bar{h}$:	(46) $\bar{l}k\bar{h}$:	(47) $\bar{l}\bar{k}h$:	(48) lkh:

$Pn\bar{3}m$ Origin 2 No. 224 (298)

(1) hkl:	(2) $\bar{h}k\bar{l}$: −110/2	(3) $\bar{h}\bar{k}l$: −101/2	(4) $h\bar{k}\bar{l}$: −011/2
(5) klh:	(6) $\bar{k}l\bar{h}$: −011/2	(7) $\bar{k}\bar{l}h$: −110/2	(8) $k\bar{l}\bar{h}$: −101/2
(9) lhk:	(10) $\bar{l}hk$: −101/2	(11) $\bar{l}\bar{h}k$: −011/2	(12) $l\bar{h}\bar{k}$: −110/2
(13) $kh\bar{l}$: −110/2	(14) $\bar{k}h\bar{l}$:	(15) $\bar{k}\bar{h}\bar{l}$: −101/2	(16) $k\bar{h}\bar{l}$: −011/2
(17) $hl\bar{k}$: −110/2	(18) $\bar{h}l\bar{k}$: −011/2	(19) $\bar{h}\bar{l}\bar{k}$:	(20) $h\bar{l}\bar{k}$: −101/2
(21) $\bar{l}kh$: −110/2	(22) $l\bar{k}h$: −101/2	(23) $lk\bar{h}$: −011/2	(24) $\bar{l}\bar{k}\bar{h}$:

$Fm\bar{3}m$ No. 225 (299)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) $kh\bar{l}$:	(14) $\bar{k}h\bar{l}$:	(15) $\bar{k}\bar{h}\bar{l}$:	(16) $k\bar{h}\bar{l}$:
(17) $hl\bar{k}$:	(18) $\bar{h}l\bar{k}$:	(19) $\bar{h}\bar{l}\bar{k}$:	(20) $h\bar{l}\bar{k}$:
(21) $\bar{l}kh$:	(22) $l\bar{k}h$:	(23) $lk\bar{h}$:	(24) $\bar{l}\bar{k}\bar{h}$:

$Fm\bar{3}c$ No. 226 (300)

(1) hkl:	(2) $\bar{h}k\bar{l}$:	(3) $\bar{h}\bar{k}l$:	(4) $h\bar{k}\bar{l}$:
(5) klh:	(6) $\bar{k}l\bar{h}$:	(7) $\bar{k}\bar{l}h$:	(8) $k\bar{l}\bar{h}$:
(9) lhk:	(10) $\bar{l}hk$:	(11) $\bar{l}\bar{h}k$:	(12) $l\bar{h}\bar{k}$:
(13) $kh\bar{l}$: −111/2	(14) $\bar{k}h\bar{l}$: −111/2	(15) $\bar{k}\bar{h}\bar{l}$: −111/2	(16) $k\bar{h}\bar{l}$: −111/2
(17) $hl\bar{k}$: −111/2	(18) $\bar{h}l\bar{k}$: −111/2	(19) $\bar{h}\bar{l}\bar{k}$: −111/2	(20) $h\bar{l}\bar{k}$: −111/2
(21) $\bar{l}kh$: −111/2	(22) $l\bar{k}h$: −111/2	(23) $lk\bar{h}$: −111/2	(24) $\bar{l}\bar{k}\bar{h}$: −111/2

$Fd\bar{3}m$ Origin 1 No. 227 (301)

(1) hkl:	(2) $\bar{h}k\bar{l}$: −011/2	(3) $\bar{h}\bar{k}l$: −110/2	(4) $h\bar{k}\bar{l}$: −101/2
(5) klh:	(6) $\bar{k}l\bar{h}$: −101/2	(7) $\bar{k}\bar{l}h$: −011/2	(8) $k\bar{l}\bar{h}$: −110/2

Table A1.4.4.1 (*cont.*)

(9) lhk:	(10) $\overline{l}hk$: −110/2	(11) $l\overline{h}k$: −101/2	(12) $\overline{l}\overline{h}k$: −011/2
(13) $kh\overline{l}$: −313/4	(14) $\overline{k}h\overline{l}$: −111/4	(15) $\overline{k}hl$: −133/4	(16) $k\overline{h}l$: −331/4
(17) $h\overline{l}k$: −313/4	(18) $\overline{h}lk$: −331/4	(19) $\overline{h}\overline{l}k$: −111/4	(20) $hl\overline{k}$: −133/4
(21) $\overline{l}kh$: −313/4	(22) $l\overline{k}h$: −133/4	(23) $lk\overline{h}$: −331/4	(24) $\overline{l}\overline{k}h$: −111/4
(25) $\overline{h}k\overline{l}$: −111/4	(26) $hk\overline{l}$: −133/4	(27) $h\overline{k}l$: −331/4	(28) $\overline{h}kl$: −313/4
(29) $k\overline{l}h$: −111/4	(30) $kl\overline{h}$: −313/4	(31) $k\overline{l}h$: −133/4	(32) $\overline{k}lh$: −331/4
(33) $\overline{l}h\overline{k}$: −111/4	(34) $lh\overline{k}$: −331/4	(35) $l\overline{h}k$: −313/4	(36) $\overline{l}hk$: −133/4
(37) $\overline{k}\overline{h}l$: −101/2	(38) khl:	(39) $k\overline{h}\overline{l}$: −011/2	(40) $\overline{k}h\overline{l}$: −110/2
(41) $\overline{h}l\overline{k}$: −101/2	(42) $h\overline{l}\overline{k}$: −110/2	(43) hlk:	(44) $\overline{h}lk$: −011/2
(45) $l\overline{k}h$: −101/2	(46) $\overline{l}k\overline{h}$: −011/2	(47) $\overline{l}\overline{k}h$: −110/2	(48) lkh:

$Fd\overline{3}m$ Origin 2 No. 227 (302)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −312/4	(3) $\overline{h}k\overline{l}$: −123/4	(4) $h\overline{k}\overline{l}$: −231/4
(5) klh:	(6) $\overline{k}lh$: −231/4	(7) $\overline{k}l\overline{h}$: −312/4	(8) $k\overline{l}\overline{h}$: −123/4
(9) lhk:	(10) $\overline{l}hk$: −123/4	(11) $\overline{l}h\overline{k}$: −231/4	(12) $l\overline{h}\overline{k}$: −312/4
(13) $kh\overline{l}$: −312/4	(14) $\overline{k}\overline{h}l$:	(15) $\overline{k}hl$: −123/4	(16) $k\overline{h}l$: −231/4
(17) $h\overline{l}k$: −312/4	(18) $\overline{h}lk$: −231/4	(19) $\overline{h}\overline{l}k$:	(20) $hl\overline{k}$: −123/4
(21) $\overline{l}kh$: −312/4	(22) $l\overline{k}h$: −123/4	(23) $lk\overline{h}$: −231/4	(24) $\overline{l}\overline{k}h$:

$Fd\overline{3}c$ Origin 1 No. 228 (303)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −011/2	(3) $\overline{h}k\overline{l}$: −110/2	(4) $h\overline{k}\overline{l}$: −101/2
(5) klh:	(6) $\overline{k}lh$: −101/2	(7) $\overline{k}l\overline{h}$: −011/2	(8) $k\overline{l}\overline{h}$: −110/2
(9) lhk:	(10) $\overline{l}hk$: −110/2	(11) $\overline{l}h\overline{k}$: −101/2	(12) $l\overline{h}\overline{k}$: −011/2
(13) $kh\overline{l}$: −313/4	(14) $\overline{k}\overline{h}l$: −111/4	(15) $\overline{k}hl$: −133/4	(16) $k\overline{h}l$: −331/4
(17) $h\overline{l}k$: −313/4	(18) $\overline{h}lk$: −331/4	(19) $\overline{h}\overline{l}k$: −111/4	(20) $hl\overline{k}$: −133/4
(21) $\overline{l}kh$: −313/4	(22) $l\overline{k}h$: −133/4	(23) $lk\overline{h}$: −331/4	(24) $\overline{l}\overline{k}h$: −111/4
(25) $\overline{h}k\overline{l}$: −333/4	(26) $hk\overline{l}$: −311/4	(27) $h\overline{k}l$: −113/4	(28) $\overline{h}kl$: −131/4
(29) $k\overline{l}h$: −333/4	(30) $kl\overline{h}$: −131/4	(31) $k\overline{l}h$: −311/4	(32) $\overline{k}lh$: −113/4
(33) $\overline{l}h\overline{k}$: −333/4	(34) $lh\overline{k}$: −113/4	(35) $l\overline{h}k$: −131/4	(36) $\overline{l}hk$: −311/4
(37) $\overline{k}\overline{h}l$: −010/2	(38) khl: −111/2	(39) $k\overline{h}\overline{l}$: −100/2	(40) $\overline{k}h\overline{l}$: −001/2
(41) $\overline{h}l\overline{k}$: −010/2	(42) $h\overline{l}\overline{k}$: −001/2	(43) hlk: −111/2	(44) $\overline{h}lk$: −100/2
(45) $l\overline{k}h$: −010/2	(46) $\overline{l}k\overline{h}$: −100/2	(47) $\overline{l}\overline{k}h$: −001/2	(48) lkh: −111/2

$Fd\overline{3}c$ Origin 2 No. 228 (304)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −132/4	(3) $\overline{h}k\overline{l}$: −321/4	(4) $h\overline{k}\overline{l}$: −213/4
(5) klh:	(6) $\overline{k}lh$: −213/4	(7) $\overline{k}l\overline{h}$: −132/4	(8) $k\overline{l}\overline{h}$: −321/4
(9) lhk:	(10) $\overline{l}hk$: −321/4	(11) $\overline{l}h\overline{k}$: −213/4	(12) $l\overline{h}\overline{k}$: −132/4
(13) $kh\overline{l}$: −310/4	(14) $\overline{k}\overline{h}l$: −111/2	(15) $\overline{k}hl$: −103/4	(16) $k\overline{h}l$: −031/4
(17) $h\overline{l}k$: −310/4	(18) $\overline{h}lk$: −031/4	(19) $\overline{h}\overline{l}k$: −111/2	(20) $hl\overline{k}$: −103/4
(21) $\overline{l}kh$: −310/4	(22) $l\overline{k}h$: −103/4	(23) $lk\overline{h}$: −031/4	(24) $\overline{l}\overline{k}h$: −111/2

$Im\overline{3}m$ No. 229 (305)

(1) hkl:	(2) $\overline{h}k\overline{l}$:	(3) $\overline{h}k\overline{l}$:	(4) $h\overline{k}\overline{l}$:
(5) klh:	(6) $\overline{k}lh$:	(7) $\overline{k}l\overline{h}$:	(8) $k\overline{l}\overline{h}$:
(9) lhk:	(10) $\overline{l}hk$:	(11) $\overline{l}h\overline{k}$:	(12) $l\overline{h}\overline{k}$:
(13) $kh\overline{l}$:	(14) $\overline{k}h\overline{l}$:	(15) $\overline{k}hl$:	(16) $k\overline{h}l$:
(17) $h\overline{l}k$:	(18) $\overline{h}lk$:	(19) $\overline{h}\overline{l}k$:	(20) $hl\overline{k}$:
(21) $\overline{l}kh$:	(22) $l\overline{k}h$:	(23) $lk\overline{h}$:	(24) $\overline{l}\overline{k}h$:

$Ia\overline{3}d$ No. 230 (306)

(1) hkl:	(2) $\overline{h}k\overline{l}$: −101/2	(3) $\overline{h}k\overline{l}$: −011/2	(4) $h\overline{k}\overline{l}$: −110/2
(5) klh:	(6) $\overline{k}lh$: −110/2	(7) $\overline{k}l\overline{h}$: −101/2	(8) $k\overline{l}\overline{h}$: −011/2
(9) lhk:	(10) $\overline{l}hk$: −011/2	(11) $\overline{l}h\overline{k}$: −110/2	(12) $l\overline{h}\overline{k}$: −101/2
(13) $kh\overline{l}$: −311/4	(14) $\overline{k}h\overline{l}$: −333/4	(15) $\overline{k}hl$: −113/4	(16) $k\overline{h}l$: −131/4
(17) $h\overline{l}k$: −311/4	(18) $\overline{h}lk$: −131/4	(19) $\overline{h}\overline{l}k$: −333/4	(20) $hl\overline{k}$: −113/4
(21) $\overline{l}kh$: −311/4	(22) $l\overline{k}h$: −113/4	(23) $lk\overline{h}$: −131/4	(24) $\overline{l}\overline{k}h$: −333/4

References

Altermatt, U. D. & Brown, I. D. (1987). *A real-space computer-based symmetry algebra. Acta Cryst.* A**43**, 125–130.

Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. (2006). *Bilbao Crystallographic Server. II. Representations of crystallographic point groups and space groups. Acta Cryst.* A**62**, 115–128.

Authier, A. (1981). *The Reciprocal Lattice*. Edited by the IUCr Commission on Crystallographic Teaching. Cardiff: University College Cardiff Press.

Bertaut, E. F. (1964). *On the symmetry of phases in the reciprocal lattice: a simple method. Acta Cryst.* **17**, 778–779.

Bienenstock, A. & Ewald, P. P. (1962). *Symmetry of Fourier space. Acta Cryst.* **15**, 1253–1261.

Bourne, P. E., Berman, H. M., McMahon, B., Watenpaugh, K. D., Westbrook, J. D. & Fitzgerald, P. M. D. (1997). *mmCIF: macromolecular crystallographic information file. Methods Enzymol.* **277**, 571–590.

Bragg, L. (1966). *The Crystalline State*, Vol. I, *A General Survey*. London: Bell.

Buerger, M. J. (1942). *X-ray Crystallography*. New York: John Wiley.

Buerger, M. J. (1949). *Crystallographic symmetry in reciprocal space. Proc. Natl Acad. Sci. USA*, **35**, 198–201.

Buerger, M. J. (1960). *Crystal Structure Analysis*. New York: John Wiley.

Computers in the New Laboratory – a Nature Survey (1981). *Nature (London)*, **290**, 193–200.

Dowty, E. (2000). *ATOMS: a program for the display of atomic structures.* Commercial software available from http://www.shapesoftware.com.

Ewald, P. P. (1921). *Das 'reziproke Gitter' in der Strukturtheorie. Z. Kristallogr. Teil A*, **56**, 129–156.

Grosse-Kunstleve, R. W. (1995). *SGINFO: a comprehensive collection of ANSI C routines for the handling of space-group symmetry*. Freely available from http://cci.lbl.gov/sginfo/.

Grosse-Kunstleve, R. W. (1999). *Algorithms for deriving crystallographic space-group information. Acta Cryst.* A**55**, 383–395.

Hall, S. R. (1981*a*). *Space-group notation with an explicit origin. Acta Cryst.* A**37**, 517–525.

Hall, S. R. (1981*b*). *Space-group notation with an explicit origin: erratum. Acta Cryst.* A**37**, 921.

Hall, S. R. (1997). *LoopX: a script used to loop Xtal.* Copyright University of Western Australia. Freely available from http://www.crystal.uwa.edu.au/~syd/symmetry/.

Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (2000). *Xtal: a system of crystallographic programs.* Copyright University of Western Australia. Freely available from http://xtal.crystal.uwa.edu.au.

Hearn, A. C. (1973). *REDUCE 2.0 user's manual*. University of Utah, Salt Lake City, Utah, USA.

Hovmöller, S. (1992). *CRISP: crystallographic image processing on a personal computer. Ultramicroscopy*, **41**, 121–135.

International Tables for Crystallography (1983). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn. Dordrecht: Reidel.

International Tables for Crystallography (1987). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, 2nd ed. Dordrecht: Reidel.

International Tables for Crystallography (1992). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, 3rd ed. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (1995). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, 4th ed. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (2002). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, 5th ed. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (2005). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, 5th ed., corrected reprint. Heidelberg: Springer.

International Tables for Crystallography (1993). Vol. B, *Reciprocal Space*, edited by U. Shmueli, 1st ed. Dordrecht: Kluwer Academic Publishers.

International Tables for X-ray Crystallography (1952). Vol. I, *Symmetry Groups*, edited by N. F. M. Henry & K. Lonsdale. Birmingham: Kynoch Press.

International Tables for X-ray Crystallography (1965). Vol. I, *Symmetry Groups*, edited by N. F. M. Henry & K. Lonsdale, 2nd ed. Birmingham: Kynoch Press.

Internationale Tabellen zur Bestimmung von Kristallstrukturen (1935). 1. Band. Edited by C. Hermann. Berlin: Borntraeger.

Koch, E. & Fischer, W. (2005). In *International Tables for Crystallography*, Vol. A, Part 11, *Symmetry operations*. Heidelberg: Springer.

Larine, M., Klimkovich, S., Farrants, G., Hovmöller, S. & Xiaodong, Z. (1995). *Space Group Explorer: a computer program for obtaining extensive information about any of the 230 space groups.* Freely available from http://www.calidris-em.com/.

Lipson, H. & Cochran, W. (1966). *The Crystalline State*, Vol. II, *The Determination of Crystal Structures*. London: Bell.

Lipson, H. & Taylor, C. A. (1958). *Fourier Transforms and X-ray Diffraction*. London: Bell.

Seitz, F. (1935). *A matrix-algebraic development of crystallographic groups. III. Z. Kristallogr.* **90**, 289–313.

Shmueli, U. (1984). *Space-group algorithms. I. The space group and its symmetry elements. Acta Cryst.* A**40**, 559–567.

Waser, J. (1955). *Symmetry relations between structure factors. Acta Cryst.* **8**, 595.

Wells, M. (1965). *Computational aspects of space-group symmetry. Acta Cryst.* **19**, 173–179.

Zachariasen, W. H. (1945). *Theory of X-ray Diffraction in Crystals*. New York: John Wiley.

International Tables for Crystallography (2010). Vol. B, Chapter 1.5, pp. 175–192.

1.5. Crystallographic viewpoints in the classification of space-group representations

By M. I. Aroyo and H. Wondratschek

1.5.1. List of abbreviations and symbols

BC	Bradley & Cracknell (1972)
CDML	Cracknell, Davies, Miller & Love (1979)
IT A	*International Tables for Crystallography*, Volume A (2005)
irreps	Irreducible representations
L; $\mathbf{L}^*$	Vector lattice of a space group $\mathcal{G}$; reciprocal lattice of $\mathcal{G}$
t	Vector of the lattice **L** of $\mathcal{G}$
k	Vector of the reciprocal space
K	Vector of the reciprocal lattice $\mathbf{L}^*$ [see Note (1)]
a, **b**, **c**; $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$	Basis vectors of the crystal lattice; basis vectors of the reciprocal lattice
$(\mathbf{a})^T$	Row of basis vectors [see Note (2)]
$(\mathbf{a}^*)$	Column of basis vectors of the reciprocal lattice $\mathbf{L}^*$
X	Point of point space
x, y, z; k_x, k_y, k_z	Point coordinates; vector coefficients
$\boldsymbol{x}$; $\boldsymbol{r}$	Column of point coordinates; column of vector coefficients
$(\boldsymbol{k})^T$	Row of coefficients of a reciprocal-space vector [see Note (2)]
a, b, c	Lengths of the basis vectors of the lattice
α, β, γ	Parameters of **k**-vector coefficients in CDML
a^*, b^*, c^*	Lengths of the basis vectors of the reciprocal lattice
$\boldsymbol{M}$, $\boldsymbol{R}$, $\boldsymbol{D}$, $\boldsymbol{S}$	Matrices
$\boldsymbol{W}$	Matrix part of a mapping
$\boldsymbol{w}$	Column part of a mapping
$(\boldsymbol{A}, \boldsymbol{a})$, $(\boldsymbol{W}, \boldsymbol{w})$	Matrix–column pairs
$\mathcal{G}$; $\mathcal{G}_0$; $(\mathcal{G})^*$	Group or space group; symmorphic space group; reciprocal-space group
$\mathcal{T}$	Translation subgroup of $\mathcal{G}$
$\mathcal{P}$ or $\overline{\mathcal{G}}$; $\mathcal{Q}$	Point group; holohedral point group
$\mathcal{S}$	Site-symmetry group
$\overline{\mathcal{G}}^{\mathbf{k}}$; $\mathcal{L}^{\mathbf{k}}$	Little co-group of **k**; little group of **k**
g, h, e	Group elements of $\mathcal{G}$
2, 3, m	Symmetry operations
$\Gamma(\mathcal{G})$	(Matrix) representation of the group $\mathcal{G}$

Notes: (1) In crystallography, vectors are designated by lower-case bold-face letters. With **K** we make an exception in order to follow the tradition of physics. A crystallographic alternative could be $\mathbf{t}^*$. (2) In crystallography, point coordinates or vector coefficients are written as columns. Therefore, columns are taken as 'normal'. In order to distinguish rows from columns (the coefficients k_i of vectors in reciprocal space, *i.e.* the Miller indices, and the basis of the crystal lattice are written as rows), rows are regarded as transposed columns and are thus marked by $(\ldots)^T$.

1.5.2. Introduction

This chapter on representations widens the scope of the general topics of reciprocal space treated in this volume.

Space-group representations play a growing role in physical applications of crystal symmetry. They are treated in a number of papers and books but comparison of the terms and the listed data is difficult. The main reason for this is the lack of standards in the classification and nomenclature of representations. As a result, the reader is confronted with barely comparable notations used by the different authors, see *e.g.* Stokes & Hatch (1988), Table 7.

The **k** vectors, which can be described as vectors in reciprocal space, play a decisive role in the description and classification of space-group representations. Their symmetry properties are determined by the so-called *reciprocal-space group* $(\mathcal{G})^*$ which is always isomorphic to a symmorphic space group $\mathcal{G}_0$. The different symmetries of **k** vectors correspond to the different kinds of point orbits in the symmorphic space groups $\mathcal{G}_0$. The classification of point orbits into Wyckoff positions in *International Tables for Crystallography* Volume A (*IT* A) (2005) can be used directly to classify the irreducible representations of a space group, abbreviated *irreps*; the Wyckoff positions of the symmorphic space groups $\mathcal{G}_0$ form a basis for a *natural* classification of the irreps. This was first discovered by Wintgen (1941). Similar results have been obtained independently by Raghavacharyulu (1961), who introduced the term reciprocal-space group. In this chapter a classification of irreps is provided which is based on Wintgen's idea.

Although this idea is now more than 60 years old, it has been utilized only rarely and has not yet found proper recognition in the literature and in the existing tables of space-group irreps. Slater (1962) described the correspondence between the special **k** vectors of the Brillouin zone and the Wyckoff positions of space group $Pm\overline{3}m$. Similarly, Jan (1972) compared Wyckoff positions with points of the Brillouin zone when describing the symmetry $Pm\overline{3}$ of the Fermi surface for the pyrite structure. However, the widespread tables of Miller & Love (1967), Zak *et al.* (1969), Bradley & Cracknell (1972) (abbreviated as BC), Cracknell *et al.* (1979) (abbreviated as CDML), and Kovalev (1986) have not made use of this kind of classification and its possibilities, and existing tables are unnecessarily complicated, *cf.* Boyle (1986).

In addition, historical reasons have obscured the classification of irreps and impeded their application. The first considerations of irreps dealt only with space groups of translation lattices (Bouckaert *et al.*, 1936). Later, other space groups were taken into consideration as well. Instead of treating these (lower) symmetries as such, their irreps were derived and classified by starting from the irreps of lattice space groups and proceeding to those of lower symmetry. This procedure has two consequences:

(1) those **k** vectors that are special in a lattice space group are also correspondingly listed in the low-symmetry space group even if they have lost their special properties due to the symmetry reduction;

(2) during the symmetry reduction unnecessary new symbols of **k** vectors are introduced.

The use of the reciprocal-space group $(\mathcal{G})^*$ avoids both these detours.

The relations between the special **k** vectors as listed by CDML and the Wyckoff positions of the space groups of *IT* A have been derived and displayed in figures and tables for a few space groups by Aroyo & Wondratschek (1995). The **k**-vector classification scheme based on Wintgen's (1941) reciprocal-space group approach has been applied meanwhile to all space groups. The compilation of Brillouin-zone figures and **k**-vector correlation tables for the 230 space groups constitutes the wavevector database of the Bilbao Crystallographic Server (1998), a website of crystallographic databases and programs that can be used free

Copyright © 2010 International Union of Crystallography

of charge from any computer with a web browser *via* the Internet (Aroyo, Perez-Mato *et al.*, 2006; Aroyo, Kirov *et al.*, 2006). Simple retrieval tools give direct access to the figures and tables for any space group. The wavevector database available on the server forms the background of the description and classification of the space-group irreps calculated and applied by different programs of the server.

This chapter is a modification of Chapter 1.5 of the second edition of *International Tables for Crystallography*, Volume B, published in 2001. As in the previous edition, we consider in more detail the reciprocal-space group approach and show that widely used crystallographic conventions can be adopted for the classification of space-group representations. Some basic concepts are developed in Section 1.5.3. Possible conventions are discussed in Section 1.5.4. In contrast to Chapter 1.5 in the second edition of *IT* B, the consequences and advantages of the reciprocal-space group approach are demonstrated and discussed in Section 1.5.5 using examples from the database of the Bilbao Crystallographic Server (1998).

1.5.3. Basic concepts

The aim of this section is to give a brief overview of some of the basic concepts related to groups and their representations. Its content should be of some help to readers who wish to refresh their knowledge of space groups and representations, and to familiarize themselves with the kind of description in this chapter. However, it can not serve as an introductory text for these subjects. The interested reader is referred to books dealing with space-group theory, representations of space groups and their applications in solid-state physics: see BC or Chapter 1.2 of *International Tables for Crystallography* Volume D by Janssen (2003).

1.5.3.1. Representations of finite groups

Group theory is the proper tool for studying symmetry in science. The elements of the crystallographic groups are rigid motions (isometries) with regard to performing one after another. The set of all isometries that map an object onto itself always fulfils the group postulates and is called the symmetry or the symmetry group of that object; the isometry itself is called a symmetry operation. Symmetry groups of crystals are dealt with in this chapter. In addition, groups of matrices with regard to matrix multiplication (matrix groups) are considered frequently. Such groups will sometimes be called realizations or representations of abstract groups.

Many applications of group theory to physical problems are closely related to representation theory, *cf.* Rosen (1981) and references therein. In this section, matrix representations Γ of finite groups $\mathcal{G}$ are considered. The concepts of *homomorphism* and *matrix groups* are of essential importance.

A group $\mathcal{B}$ is a homomorphic image of a group $\mathcal{A}$ if there exists a mapping of the elements a_i of $\mathcal{A}$ onto the elements b_k of $\mathcal{B}$ that preserves the multiplication relation (in general several elements of $\mathcal{A}$ are mapped onto one element of $\mathcal{B}$): if $\mathsf{a}_i \rightarrow \mathsf{b}_i$ and $\mathsf{a}_k \rightarrow \mathsf{b}_k$, then $\mathsf{a}_i\mathsf{a}_k \rightarrow \mathsf{b}_i\mathsf{b}_k$ holds for all elements of $\mathcal{A}$ and $\mathcal{B}$ (the image of the product is equal to the product of the images). In the special case of a one-to-one mapping, the homomorphism is called an *isomorphism*.

A matrix group is a group whose elements are non-singular square matrices. The law of combination is matrix multiplication and the group inverse is the inverse matrix. In the following we will be concerned with some basic properties of finite matrix groups relevant to representations.

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two matrix groups whose matrices are of the same dimension. They are said to be equivalent if there exists a (non-singular) matrix $\boldsymbol{S}$ such that $\mathcal{M}_2 = \boldsymbol{S}^{-1}\mathcal{M}_1\boldsymbol{S}$ holds. Equivalence implies isomorphism but the inverse is not true: two matrix groups may be isomorphic without being equivalent. According to the theorem of Schur-Auerbach, every finite matrix group is equivalent to a unitary matrix group (by a unitary matrix group we understand a matrix group consisting entirely of unitary matrices).

A matrix group $\mathcal{M}$ is *reducible* if it is equivalent to a matrix group in which every matrix $\boldsymbol{M}$ is of the form

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{X} \\ \boldsymbol{O} & \boldsymbol{D}_2 \end{pmatrix},$$

see *e.g.* Lomont (1959), p. 47. The group $\mathcal{M}$ is *completely reducible* if it is equivalent to a matrix group in which for all matrices $\boldsymbol{R}$ the submatrices $\boldsymbol{X}$ are $\boldsymbol{O}$ matrices (consisting of zeros only). According to the theorem of Maschke, a finite matrix group is completely reducible if it is reducible. A matrix group is *irreducible* if it is not reducible.

A (matrix) representation $\Gamma(\mathcal{G})$ of a group $\mathcal{G}$ is a homomorphic mapping of $\mathcal{G}$ onto a matrix group $\mathcal{M}(\mathcal{G})$. In a representation Γ every element $\mathsf{g} \in \mathcal{G}$ is associated with a matrix $\boldsymbol{M}(\mathsf{g})$. The dimension of the matrices is called the dimension of the representation.

The above-mentioned theorems on finite matrix groups can be applied directly to representations: we can restrict the considerations to unitary representations only. Further, since every finite matrix group is either completely reducible into irreducible constituents or irreducible, it follows that the infinite set of all matrix representations of a group is known in principle once the irreps are known. Naturally, the question of how to construct all nonequivalent irreps of a finite group and how to classify them arises.

Linear representations are especially important for applications. In this chapter only linear representations of space groups will be considered. Realizations and representations are homomorphic images of abstract groups, but not all of them are linear. In particular, the action of space groups on point space is a nonlinear realization of the abstract space groups because isometries and thus symmetry operations of space groups $\mathcal{G}$ are nonlinear operations. The same holds for their description by matrix-column pairs $(\boldsymbol{W}, \boldsymbol{w})$,[1] by the general position, or by augmented (4×4) matrices, see *IT* A, Part 8. Therefore, the isomorphic matrix representation of a space group, mostly used by crystallographers and listed in the space-group tables of *IT* A as the general position, is not linear.

1.5.3.2. Space groups

In crystallography one deals with real crystals. In many cases the treatment of the crystal is much simpler, but nevertheless describes the crystal and its properties very well, if the real crystal is replaced by an 'ideal crystal'. The real crystal is then considered to be a finite piece of an undisturbed, periodic, and thus infinitely extended arrangement of particles or their centres: ideal crystals are periodic objects in three-dimensional point space E^3, also called direct space. Periodicity means that there are translations among the symmetry operations of ideal crystals. The symmetry group of an ideal crystal is called its space group $\mathcal{G}$.

Space groups $\mathcal{G}$ are of special interest for our problem because:

(1) their irreps are the subject of the classification to be discussed;

(2) this classification makes use of the isomorphism of certain groups to the so-called symmorphic space groups $\mathcal{G}_0$.

Therefore, space groups are introduced here in a slightly more detailed manner than the other concepts. In doing this we follow the definitions and symbolism of *IT* A, Part 8.

[1] In physics written as the Seitz symbol $(\boldsymbol{W}|\boldsymbol{w})$.

To each space group $\mathcal{G}$ belongs an infinite set $\mathcal{T}$ of translations, the *translation subgroup* of $\mathcal{G}$. The group $\mathcal{T}$ forms an infinite Abelian invariant subgroup of $\mathcal{G}$. For each translation its translation vector is defined. The set of all translation vectors is called the *vector lattice* $\mathbf{L}$ of $\mathcal{G}$. Because of the finite size of the atoms constituting the real crystal, the lengths of the translation vectors of the ideal crystal cannot be arbitrarily small; rather there is a lower limit $\delta > 0$ for their length in the range of a few Å.

When referred to a coordinate system $(O, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, consisting of an origin O and a basis $\mathbf{a}_k$, the symmetry operations, *i.e.* the elements $\mathsf{g}: \tilde{\boldsymbol{x}} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{w}$ of the space group $\mathcal{G}$, are described by matrix–column pairs $(\boldsymbol{W}, \boldsymbol{w})$ with matrix part $\boldsymbol{W}$ and column part $\boldsymbol{w}$. The translations of $\mathcal{G}$ are represented by pairs $(\boldsymbol{I}, \boldsymbol{t}_i)$, where $\boldsymbol{I}$ is the (3×3) unit matrix and $\boldsymbol{t}_i$ is the column of coefficients of the translation vector $\mathbf{t}_i \in \mathbf{L}$. The basis can always be chosen such that all columns $\boldsymbol{t}_i$ and no other columns of translations consist of integers. Such a basis $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ is called a *primitive basis*. For each vector lattice $\mathbf{L}$ there exists an infinite number of primitive bases.

The space group $\mathcal{G}$ can be decomposed into right cosets relative to $\mathcal{T}$:

$$\mathcal{G} = \mathcal{T} \cup \mathcal{T}(\boldsymbol{W}_2, \boldsymbol{w}_2) \cup \ldots \cup \mathcal{T}(\boldsymbol{W}_i, \boldsymbol{w}_i) \cup \ldots \cup \mathcal{T}(\boldsymbol{W}_n, \boldsymbol{w}_n). \quad (1.5.3.1)$$

The coset representatives form the finite set $\mathrm{V} = \{(\boldsymbol{W}_v, \boldsymbol{w}_v)\}$, $v = 1, \ldots, n$, with $(\boldsymbol{W}_1, \boldsymbol{w}_1) = (\boldsymbol{I}, \boldsymbol{o})$, where $\boldsymbol{I}$ is the unit matrix and $\boldsymbol{o}$ is the column consisting of zeros only. The factor group $\mathcal{G}/\mathcal{T}$ is isomorphic to the *point group* $\mathcal{P}$ of $\mathcal{G}$ (called $\overline{\mathcal{G}}$ in books on representation theory) describing the symmetry of the external shape of the macroscopic crystal and being represented by the matrices $\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_n$, $\mathcal{P} = \{\boldsymbol{W}_i\}$. If V can be chosen such that all $\boldsymbol{w}_v = \boldsymbol{o}$, then $\mathcal{G}$ is called a *symmorphic space group* $\mathcal{G}_0$. A symmorphic space group can be recognized easily from its conventional Hermann–Mauguin symbol which does not contain any screw or glide component. In terms of group theory, a symmorphic space group is the semidirect product of $\mathcal{T}$ and $\mathcal{P}$, *cf.* BC, p. 44. In symmorphic space groups $\mathcal{G}_0$ (and in no others) there are site-symmetry groups which are isomorphic to the point group $\mathcal{P}$ of $\mathcal{G}_0$. (For the term 'site-symmetry group', see Section 1.5.4.2.)

Space groups can be classified into 219 (*affine*) *space-group types* either by isomorphism or by affine equivalence; the 230 *crystallographic* space-group types are obtained by restricting the transformations available for affine equivalence to those with positive determinant, *cf. IT* A, Section 8.2.1. Many important properties of space groups are shared by all space groups of a type. In such a case one speaks of *properties of the type.* For example, if a space group is symmorphic, then all space groups of its type are symmorphic, so that one normally speaks of a symmorphic space-group type.

With the concept of symmorphic space groups one can also define the arithmetic crystal classes: Let $\mathcal{G}_0$ be a symmorphic space group referred to a primitive basis and $\mathrm{V} = \{(\boldsymbol{W}_v, \boldsymbol{w}_v)\}$ its set of coset representatives with $\boldsymbol{w}_v = \boldsymbol{o}$ for all columns. To $\mathcal{G}_0$ all those space groups $\mathcal{G}$ can be assigned for which a primitive basis can be found such that the matrix parts $\boldsymbol{W}_v$ of their sets V are the same as those of $\mathcal{G}_0$, only the columns $\boldsymbol{w}_v$ may differ. In this way, to a type of symmorphic space groups $\mathcal{G}_0$, other types of space groups are assigned, *i.e.* the space-group types are classified according to the symmorphic space-group types. These classes are called *arithmetic crystal classes* of space groups or of space-group types.

There are 73 arithmetic crystal classes corresponding to the 73 types of symmorphic space groups; between 1 and 16 space-group types belong to an arithmetic crystal class. A matrix-algebraic definition of arithmetic crystal classes and a proposal for their nomenclature can be found in *IT* A, Section 8.2.3; see also Section 8.3.4 and Table 8.3.4.1.

1.5.3.3. Representations of the translation group $\mathcal{T}$ and the reciprocal lattice

For representation theory we follow the terminology of BC and CDML.

Let $\mathcal{G}$ be referred to a primitive basis. For the following, the infinite set of translations, based on discrete cyclic groups of infinite order, will be replaced by a (very large) finite set in the usual way, assuming the Born–von Karman boundary conditions

$$(\boldsymbol{I}, \boldsymbol{t}_{bi})^{N_i} = (\boldsymbol{I}, N_i) = (\boldsymbol{I}, \boldsymbol{o}) \quad (1.5.3.2)$$

to hold, where $\boldsymbol{t}_{bi} = (1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$ and N_i is a large integer for $i = 1$, 2 or 3, respectively. Then for any lattice translation $(\boldsymbol{I}, \boldsymbol{t})$,

$$(\boldsymbol{I}, \boldsymbol{N}\boldsymbol{t}) = (\boldsymbol{I}, \boldsymbol{o}) \quad (1.5.3.3)$$

holds, where $\boldsymbol{N}\boldsymbol{t}$ is the column $(N_1 t_1, N_2 t_2, N_3 t_3)$. If the (infinitely many) translations mapped in this way onto $(\boldsymbol{I}, \boldsymbol{o})$ form a normal subgroup $\mathcal{T}_1$ of $\mathcal{G}$, then the mapping described by (1.5.3.3) is a homomorphism. There exists a factor group $\mathcal{G}' = \mathcal{G}/\mathcal{T}_1$ of $\mathcal{G}$ relative to $\mathcal{T}_1$ with translation subgroup $\mathcal{T}' = \mathcal{T}/\mathcal{T}_1$ which is finite and is sometimes called the *finite space group*.

Only the irreps of these finite space groups will be considered. The definitions of space-group type, symmorphic space group *etc.* can be transferred to these groups. Because $\mathcal{T}$ is Abelian, $\mathcal{T}'$ is also Abelian. Replacing the space group $\mathcal{G}$ by $\mathcal{G}'$ means that the especially well developed theory of representations of finite groups can be applied, *cf.* Lomont (1959), Jansen & Boon (1967). For convenience, the prime $'$ will be omitted and the symbol $\mathcal{G}$ will be used instead of $\mathcal{G}'$; $\mathcal{T}'$ will be denoted by $\mathcal{T}$ in the following.

Because $\mathcal{T}$ (formerly $\mathcal{T}'$) is Abelian, its irreps $\Gamma(\mathcal{T})$ are one-dimensional and consist of (complex) roots of unity. Owing to equations (1.5.3.2) and (1.5.3.3), the irreps $\Gamma^{q_1 q_2 q_3}[(\boldsymbol{I}, \boldsymbol{t})]$ of $\mathcal{T}$ have the form

$$\Gamma^{q_1 q_2 q_3}[(\boldsymbol{I}, \boldsymbol{t})] = \exp\left[-2\pi i\left(q_1 \frac{t_1}{N_1} + q_2 \frac{t_2}{N_2} + q_3 \frac{t_3}{N_3}\right)\right], \quad (1.5.3.4)$$

where $\boldsymbol{t}$ is the column (t_1, t_2, t_3), $q_j = 0, 1, 2, \ldots, N_j - 1$, $j = 1, 2, 3$, and t_k and q_j are integers.

Given a primitive basis $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ of $\mathbf{L}$, mathematicians and crystallographers define the *basis of the dual or reciprocal lattice* $\mathbf{L}^*$ by

$$\mathbf{a}_i \cdot \mathbf{a}_j^* = \delta_{ij}, \quad (1.5.3.5)$$

where $\mathbf{a} \cdot \mathbf{a}^*$ is the scalar product between the vectors and δ_{ij} is the unit matrix (see *e.g.* Chapter 1.1, Section 1.1.3). Texts on the physics of solids redefine the basis $\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*$ of the *reciprocal lattice* $\mathbf{L}^*$, lengthening each of the basis vectors $\mathbf{a}_j^*$ by the factor 2π. Therefore, in the physicist's convention the relation between the bases of direct and reciprocal lattice reads (*cf.* BC, p. 86):

$$\mathbf{a}_i \cdot \mathbf{a}_j^* = 2\pi\delta_{ij}. \quad (1.5.3.6)$$

In the present chapter only the physicist's basis of the reciprocal lattice is employed, and hence the use of $\mathbf{a}_i^*$ should not lead to misunderstandings. The set of all vectors $\mathbf{K}$,[2]

$$\mathbf{K} = k_1\mathbf{a}_1^* + k_2\mathbf{a}_2^* + k_3\mathbf{a}_3^*, \quad (1.5.3.7)$$

k_i integer, is called the lattice reciprocal to $\mathbf{L}$ or the *reciprocal lattice* $\mathbf{L}^*$.[3]

If one adopts the notation of *IT* A, Part 5, the basis of direct space is denoted by a row $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)^T$, where $(\,)^T$ means trans-

[2] See Note (1) in Section 1.5.1.

[3] The lattice $\mathbf{L}$ is often called the *direct lattice*. These names are historically introduced and cannot be changed, although equations (1.5.3.5) and (1.5.3.6) show that essentially neither of the lattices is preferred: they form a pair of *mutually reciprocal* lattices.

posed. For reciprocal space, the basis is described by a column $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$.

To each lattice generated from a basis $(\mathbf{a}_i)^T$ a reciprocal lattice is generated from the basis $(\mathbf{a}_j^*)$. Both lattices, $\mathbf{L}$ and $\mathbf{L}^*$, can be compared most easily by referring the direct lattice $\mathbf{L}$ to its *conventional* basis $(\mathbf{a}_i)^T$ as defined in Chapters 2.1 and 9.1 of *IT* A. In this case, the lattice $\mathbf{L}$ may be primitive or centred. If $(\mathbf{a}_i)^T$ forms a primitive basis of $\mathbf{L}$, *i.e.* if $\mathbf{L}$ is primitive, then the basis $(\mathbf{a}_j^*)$ forms a primitive basis of $\mathbf{L}^*$. If $\mathbf{L}$ is centred, *i.e.* $(\mathbf{a}_i)^T$ is not a primitive basis of $\mathbf{L}$, then there exists a centring matrix $\boldsymbol{P}$, $0 < \det(\boldsymbol{P}) < 1$, by which three linearly independent vectors of $\mathbf{L}$ with rational coefficients are generated from those with integer coefficients, *cf. IT* A, Table 5.1.3.1.

Moreover, $\boldsymbol{P}$ can be chosen such that the set of vectors

$$(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)^T = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)^T \boldsymbol{P} \qquad (1.5.3.8)$$

forms a primitive basis of $\mathbf{L}$. Then the basis vectors $(\mathbf{p}_1^*, \mathbf{p}_2^*, \mathbf{p}_3^*)$ of the lattice reciprocal to the lattice generated by $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)^T$ are determined by

$$(\mathbf{p}_1^*, \mathbf{p}_2^*, \mathbf{p}_3^*) = \boldsymbol{P}^{-1}(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*) \qquad (1.5.3.9)$$

and form a primitive basis of $\mathbf{L}^*$.

Because $\det(\boldsymbol{P}^{-1}) > 1$, not all vectors $\mathbf{K}$ of the form (1.5.3.7) belong to $\mathbf{L}^*$. If k_1, k_2, k_3 are the (integer) coefficients of these vectors $\mathbf{K}$ referred to $(\mathbf{a}_j^*)$ and $k_{p1}\mathbf{p}_1^* + k_{p2}\mathbf{p}_2^* + k_{p3}\mathbf{p}_3^*$ are the vectors of $\mathbf{L}^*$, then $\mathbf{K} = (k_j)^T(\mathbf{a}_j^*) = (k_j)^T\boldsymbol{P}(\mathbf{p}_i^*) = (k_{pi})^T(\mathbf{p}_i^*)$ is a vector of $\mathbf{L}^*$ if and only if the coefficients

$$(k_{p1}, k_{p2}, k_{p3})^T = (k_1, k_2, k_3)^T \boldsymbol{P} \qquad (1.5.3.10)$$

are integers. In other words, $(k_1, k_2, k_3)^T$ has to fulfil the equation

$$(k_1, k_2, k_3)^T = (k_{p1}, k_{p2}, k_{p3})^T \boldsymbol{P}^{-1}. \qquad (1.5.3.11)$$

As is well known, the Bravais type of the reciprocal lattice $\mathbf{L}^*$ is not necessarily the same as that of its direct lattice $\mathbf{L}$. If $\boldsymbol{W}$ is the matrix of a (point-) symmetry operation of the direct lattice, referred to its basis $(\mathbf{a}_i)^T$, then $\boldsymbol{W}^{-1}$ is the matrix of the same symmetry operation of the reciprocal lattice but referred to the dual basis $(\mathbf{a}_i^*)$. This does not affect the symmetry because in a (symmetry) group the inverse of each element in the group also belongs to the group. Therefore, the (point) symmetries of a lattice and its reciprocal lattice are always the same. However, there may be differences in the matrix descriptions due to the different orientations of $\mathbf{L}$ and $\mathbf{L}^*$ relative to the symmetry elements of $\overline{\mathcal{G}}$ and due to the reference to the different bases $(\mathbf{a}_i)^T$ and $(\mathbf{a}_i^*)$. For example, if $\mathbf{L}$ has the point symmetry (Hermann–Mauguin symbol) $\overline{3}m1$, then the symbol for the point symmetry of $\mathbf{L}^*$ is $\overline{3}1m$ and *vice versa*.

1.5.3.4. Irreducible representations of space groups and the reciprocal-space group

Let $(\mathbf{a}_i)^T$ be a conventional basis of the lattice $\mathbf{L}$ of the space group $\mathcal{G}$. From (1.5.3.6), $k_i = q_i/N_i$ and $\mathbf{k} = \sum_{k=1}^{3} k_i \mathbf{a}_i^*$, equation (1.5.3.4) can be written

$$\Gamma^{q_1 q_2 q_3}[(\boldsymbol{I}, \boldsymbol{t})] = \Gamma^{\mathbf{k}}[(\boldsymbol{I}, \boldsymbol{t})] = \exp[-i\mathbf{k} \cdot \mathbf{t}]. \qquad (1.5.3.12)$$

Equation (1.5.3.12) has the same form if a primitive basis $(\mathbf{p}_i)^T$ of $\mathbf{L}$ has been chosen. In this case, the vector $\mathbf{k}$ is given by $\mathbf{k} = \sum_{i=1}^{3} k_{pi}\mathbf{p}_i^*$ with rational coefficients k_{pi}.

Let a primitive basis $(\mathbf{p}_i)^T$ be chosen for the lattice $\mathbf{L}$. The set of all vectors $\mathbf{k}$ (known as wavevectors) forms a discontinuous array. Consider two wavevectors $\mathbf{k}$ and $\mathbf{k}' = \mathbf{k} + \mathbf{K}$, where $\mathbf{K}$ is a vector of the reciprocal lattice $\mathbf{L}^*$. Obviously, $\mathbf{k}$ and $\mathbf{k}'$ describe the same irrep of $\mathcal{T}$. Therefore, to determine all irreps of $\mathcal{T}$ it is necessary to consider only the wavevectors of a small region of the reciprocal space, where the translation of this region by all vectors of $\mathbf{L}^*$ fills the reciprocal space without gap or overlap. Such a region is called a *fundamental region* of $\mathbf{L}^*$. (The nomenclature in literature is not quite uniform. We follow here widely adopted definitions.)

The fundamental region of $\mathbf{L}^*$ is not uniquely determined. Two types of fundamental regions are of interest in this chapter:

(1) The *first Brillouin zone* is that range of $\mathbf{k}$ space around $\mathbf{o}$ for which $|\mathbf{k}| \leq |\mathbf{K} - \mathbf{k}|$ holds for any vector $\mathbf{K} \in \mathbf{L}^*$ (*Wigner–Seitz cell* or *domain of influence* in $\mathbf{k}$ space). Visually, it is the region of endpoints of the $\mathbf{k}$ vectors that are nearer to the origin than to any other point of the reciprocal lattice. The Brillouin zone is a centrosymmetric body bounded by centrosymmetric planes which bisect perpendicularly the connecting lines from the origin to the neighbouring reciprocal-lattice points. The Brillouin zone is used in books and articles on irreps of space groups.

(2) The *crystallographic unit cell in reciprocal space*, for short *unit cell*, is the set of all $\mathbf{k}$ vectors with $0 \leq k_i < 1$. It corresponds to the unit cell used in crystallography for the description of crystal structures in direct space.

Let $\mathbf{k}$ be some vector according to (1.5.3.12) and $\boldsymbol{W}$ be the matrices of $\overline{\mathcal{G}}$. The following definitions are useful:

Definition. The set of all vectors $\mathbf{k}'$ fulfilling the condition

$$\mathbf{k}' = \mathbf{k}\boldsymbol{W} + \mathbf{K}, \quad \boldsymbol{W} \in \overline{\mathcal{G}}, \quad \mathbf{K} \in \mathbf{L}^* \qquad (1.5.3.13)$$

is called *the orbit of* $\mathbf{k}$.

Definition. The set of all matrices $\boldsymbol{W} \in \overline{\mathcal{G}}$ for which

$$\mathbf{k} = \mathbf{k}\boldsymbol{W} + \mathbf{K}, \quad \mathbf{K} \in \mathbf{L}^* \qquad (1.5.3.14)$$

forms a group which is called the *little co-group* $\overline{\mathcal{G}}^{\mathbf{k}}$ of $\mathbf{k}$. The vector $\mathbf{k}$ is called *general* if $\overline{\mathcal{G}}^{\mathbf{k}} = \{\boldsymbol{I}\}$; otherwise $\overline{\mathcal{G}}^{\mathbf{k}} > \{\boldsymbol{I}\}$ and $\mathbf{k}$ is called *special*.

In words: The $\mathbf{k}$ vector is called *general* if its little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$ is the identity of $\overline{\mathcal{G}}$. Otherwise, $\overline{\mathcal{G}}^{\mathbf{k}}$ is a non-trivial subgroup of the point group $\overline{\mathcal{G}}$ of $\mathcal{G}$, $\overline{\mathcal{G}} > \overline{\mathcal{G}}^{\mathbf{k}}$, and $\mathbf{k}$ is called a *special vector* of the reciprocal space.

Equation (1.5.3.14) for $\mathbf{k}$ resembles the equation

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{t}, \quad \mathbf{t} \in \mathbf{L} \qquad (1.5.3.15)$$

by which the fixed points of the symmetry operation $(\boldsymbol{W}, \boldsymbol{t})$ of a symmorphic space group $\mathcal{G}_0$ are determined. Indeed, the orbits of $\mathbf{k}$ defined by (1.5.3.13) correspond to the point orbits of $\mathcal{G}_0$, the little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$ of $\mathbf{k}$ corresponds to the site-symmetry group of that point $\boldsymbol{X}$ whose coordinates (x_i) have the same values as the vector coefficients $(k_i)^T$ of $\mathbf{k}$.

Consider the coset decomposition of $\overline{\mathcal{G}}$ relative to $\overline{\mathcal{G}}^{\mathbf{k}}$.

Definition. If $\{\boldsymbol{W}_m\}$ is a set of coset representatives of $\overline{\mathcal{G}}$ relative to $\overline{\mathcal{G}}^{\mathbf{k}}$, then the set $\{\mathbf{k}\boldsymbol{W}_m\}$ is called the *star* of $\mathbf{k}$ and the vectors $\mathbf{k}\boldsymbol{W}_m$ are called the *arms of the star*.

The number of arms of the star of $\mathbf{k}$ is equal to the order $|\overline{\mathcal{G}}|$ of the point group $\overline{\mathcal{G}}$ divided by the order $|\overline{\mathcal{G}}^{\mathbf{k}}|$ of the symmetry group $\overline{\mathcal{G}}^{\mathbf{k}}$ of $\mathbf{k}$. If $\mathbf{k}$ is general, then there are $|\overline{\mathcal{G}}|$ vectors from the orbit of $\mathbf{k}$ in each fundamental region and $|\overline{\mathcal{G}}|$ arms of the star. If $\mathbf{k}$ is special with little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$, then the number of arms of the star of $\mathbf{k}$ and the number of $\mathbf{k}$ vectors in the fundamental region from the orbit of $\mathbf{k}$ is $|\overline{\mathcal{G}}|/|\overline{\mathcal{G}}^{\mathbf{k}}|$.

Definition. The group of all elements $(\boldsymbol{W}, \boldsymbol{w}) \in \mathcal{G}$ for which $\boldsymbol{W} \in \overline{\mathcal{G}}^{\mathbf{k}}$ is called the *little group* $\mathcal{L}^{\mathbf{k}}$ *of* $\mathbf{k}$.

The analogue of the little group $\mathcal{L}^{\mathbf{k}}$ is rarely considered in crystallography.

All symmetry operations of $\mathcal{G}_0$ may be obtained as combinations of an operation that leaves the origin fixed with a transla-

tion of $\mathbf{L}$, *i.e.* are of the kind $(\boldsymbol{W}, \boldsymbol{t}) = (\boldsymbol{I}, \boldsymbol{t})(\boldsymbol{W}, \boldsymbol{o})$. We now define the analogous group for the $\mathbf{k}$ vectors. Whereas $\mathcal{G}_0$ is a realization of the corresponding abstract group in direct (point) space, the group to be defined will be a realization of it in reciprocal (vector) space.

Definition. The group $(\mathcal{G})^*$ which is the semidirect product of the point group $\overline{\mathcal{G}}$ and the translation group of the reciprocal lattice $\mathbf{L}^*$ of $\mathcal{G}$ is called the *reciprocal-space group of* $\mathcal{G}$.

By this definition, the reciprocal-space group $(\mathcal{G})^*$ is isomorphic to a symmorphic space group $\mathcal{G}_0$. The elements of $\mathcal{G}_0$ are the operations $(\boldsymbol{W}, \boldsymbol{t}_{\boldsymbol{K}}) = (\boldsymbol{I}, \boldsymbol{t}_{\boldsymbol{K}})(\boldsymbol{W}, \boldsymbol{o})$ with $\boldsymbol{W} \in \overline{\mathcal{G}}$ and $\boldsymbol{t}_{\boldsymbol{K}} \in \mathbf{L}$ with the coefficients of $\mathbf{K}$. In order to emphasize that $(\mathcal{G})^*$ is a group acting on reciprocal space and not the inverse of a space group (whatever that may mean) we insert a hyphen '-' between 'reciprocal' and 'space'.

From the definition of $(\mathcal{G})^*$ it follows that space groups of the same type define the same type of reciprocal-space group $(\mathcal{G})^*$. Moreover, as $(\mathcal{G})^*$ does not depend on the column parts of the space-group operations, all space groups of the same arithmetic crystal class determine the same type of $(\mathcal{G})^*$; for arithmetic crystal class see Section 1.5.3.2. Following Wintgen (1941), the types of reciprocal-space groups $(\mathcal{G})^*$ are listed for the arithmetic crystal classes of space groups, *i.e.* for all space groups $\mathcal{G}$, in Appendix A1.5.1.

1.5.4. Conventions in the classification of space-group irreps

Because of the isomorphism between the reciprocal-space groups $(\mathcal{G})^*$ and the symmorphic space groups $\mathcal{G}_0$ one can introduce crystallographic conventions in the classification of space-group irreps. These conventions will be compared with those which have mainly been used up to now. Illustrative examples to the following more theoretical considerations are discussed in Section 1.5.5.

1.5.4.1. Fundamental regions

Different types of regions of reciprocal space may be chosen as fundamental regions, see Section 1.5.3.4. The most frequently used type is the first Brillouin zone, which is the Wigner–Seitz cell (or Voronoi region, Dirichlet domain, domain of influence; *cf. IT* A, Chapter 9.1) of the reciprocal lattice. The Brillouin zone has the property of including the star of each $\mathbf{k}$ vector that belongs to it. Such a choice has three advantages:

(1) the Brillouin zone manifests the point symmetry of the reciprocal lattice $\mathbf{L}^*$ of $\mathcal{G}$;

(2) only $\mathbf{k}$ vectors of the boundary of the Brillouin zone may have little-group representations which are obtained from projective representations of the little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$, see *e.g.* BC, p. 156;

(3) for physical reasons, the Brillouin zone may be the most convenient fundamental region.

Of these advantages only the third may be essential. For the classification of irreps the minimal domains, see Section 1.5.4.2, are much more important than the fundamental regions. The minimal domain does not display the point-group symmetry anyway and the distinguished $\mathbf{k}$ vectors always belong to its boundary however the minimal domain may be chosen.

The serious disadvantage of the Brillouin zone is its often complicated shape which, moreover, depends on the lattice parameters of $\mathbf{L}^*$, *cf.* Section 1.5.5.3. The body that represents the Brillouin zone belongs to one of the five Fedorov polyhedra (more or less distorted versions of the cubic forms cube, rhombdodecahedron or cuboctahedron, of the hexagonal prism, or of the tetragonal elongated rhombdodecahedron). A more detailed description is that by the 24 *symmetrische Sorten* (Delaunay sorts) of Delaunay (1933*a*,*b*), Figs. 11 and 12. According to this classification, the Brillouin zone may display three types of polyhedra of cubic, one type of hexagonal, two of rhombohedral, three of tetragonal, six of orthorhombic, six of monoclinic, and three types of triclinic symmetry.

For low symmetries the shape of the Brillouin zone is so variable that BC, p. 90 *ff.* chose a primitive unit cell of $\mathbf{L}^*$ for the fundamental regions of triclinic and monoclinic crystals. This cell also reflects the point symmetry of $\mathbf{L}^*$, it has six faces only, and although its shape varies with the lattice parameters all cells are affinely equivalent. For space groups of higher symmetry, BC and most other authors prefer the Brillouin zone.

Considering $\mathbf{L}^*$ as a lattice, one can refer it to its conventional crystallographic lattice basis. Referred to this basis, the *unit cell of* $\mathbf{L}^*$ is always an alternative to the Brillouin zone. With the exception of the hexagonal lattice, the unit cell of $\mathbf{L}^*$ reflects the point symmetry, it has only six faces and its shape is always affinely equivalent for varying lattice parameters. For a space group $\mathcal{G}$ with a primitive lattice, the above-defined conventional unit cell of $\mathbf{L}^*$ is also primitive. If $\mathcal{G}$ has a centred lattice, then $\mathbf{L}^*$ also belongs to a type of centred lattice and the *conventional* cell of $\mathbf{L}^*$ [not to be confused with the cell spanned by the basis $(\mathbf{a}_j^*)$ dual to the basis $(\mathbf{a}_i)^T$] is larger than necessary. However, this is not disturbing because in this context the fundamental region is an auxiliary construction only for the definition of the minimal domain; see Section 1.5.4.2.

1.5.4.2. Minimal domains

One can show that all irreps of $\mathcal{G}$ can be built up from the irreps $\Gamma^{\mathbf{k}}$ of $\mathcal{T}$. Moreover, to find all irreps of $\mathcal{G}$ it is only necessary to consider one $\mathbf{k}$ vector from each orbit of $\mathbf{k}$, *cf.* CDML, p. 31.

Definition. A simply connected part of the fundamental region which contains *exactly one* $\mathbf{k}$ vector of each orbit of $\mathbf{k}$ is called a *minimal domain* Φ.

The choice of the minimal domain is by no means unique. One of the difficulties in comparing the published data on irreps of space groups is due to the different minimal domains found in the literature.

The number of $\mathbf{k}$ vectors of each general $\mathbf{k}$ orbit in a fundamental region is always equal to the order of the point group $\overline{\mathcal{G}}$ of $\mathcal{G}$; see Section 1.5.3.4. Therefore, the volume of the minimal domain Φ in reciprocal space is $1/|\overline{\mathcal{G}}|$ of the volume of the fundamental region. Now we can restrict the search for all irreps of $\mathcal{G}$ to the $\mathbf{k}$ vectors within a minimal domain Φ.

In general, in representation theory of space groups the Brillouin zone is taken as the fundamental region and Φ is called a *representation domain*.[4] Again, the volume of a representation domain in reciprocal space is $1/|\overline{\mathcal{G}}|$ of the volume of the Brillouin zone. In addition, as the Brillouin zone contains for each $\mathbf{k}$ vector all $\mathbf{k}$ vectors of the star of $\mathbf{k}$, by application of all symmetry operations $\boldsymbol{W} \in \overline{\mathcal{G}}$ to Φ one obtains the Brillouin zone; *cf.* BC, p. 147. As the Brillouin zone may change its geometrical type depending on the lattice parameters, the type of the representation domain may also vary with varying lattice parameters; see the examples of Section 1.5.5.

The simplest crystal structures are the lattice-like structures that are built up of translationally equivalent points (centres of particles) only. For such a structure the point group $\overline{\mathcal{G}}$ of the space group $\mathcal{G}$ is equal to the point group $\mathcal{Q}$ of its lattice $\mathbf{L}$. Such point groups are called *holohedral*, the space group $\mathcal{G}$ is called *holo-*

[4] From definition 3.7.1 on p. 147 of BC, it does not follow that a representation domain contains *exactly one* $\mathbf{k}$ vector from each star. The condition 'The intersection of the representation domain with its symmetrically equivalent domains is empty' is missing. Lines 14 to 11 from the bottom of p. 149, however, state that such a property of the representation domain is intended. The representation domains of CDML, Figs. 3.15–3.29 contain *at least one* $\mathbf{k}$ vector of each star (Vol. 1, pp. 31, 57 and 65). On pp. 66, 67 a procedure is described for eliminating those $\mathbf{k}$ vectors from the representation domain which occur more than once. In the definition of Altmann (1977), p. 204, the representation domain contains *exactly one* arm (prong) per star.

symmetric. There are seven holohedral point groups of three dimensions: $\overline{1}, 2/m, mmm, 4/mmm, \overline{3}m, 6/mmm$ and $m\overline{3}m$. For the non-holosymmetric space groups $\mathcal{G}, \overline{\mathcal{G}} < \mathcal{Q}$ holds. [The group $\mathcal{Q}$ may have *accidentally higher symmetry* in addition, *e.g.* if the lattice of a monoclinic crystal structure is orthorhombic at a certain temperature, pressure and composition. In Section 8.2.6 of *IT* A it is shown how such accidental symmetry can be treated.]

In books on representation theory of space groups, holosymmetric space groups play a distinguished role. Their representation domains are called *basic domains* Ω. For holosymmetric space groups $\Omega = \Phi$ holds. If $\mathcal{G}$ is non-holosymmetric, *i.e.* $\overline{\mathcal{G}} < \mathcal{Q}$ holds, Ω is defined by $\mathcal{Q}$ and is smaller than the representation domain Φ by a factor which is equal to the index of $\overline{\mathcal{G}}$ in $\mathcal{Q}$. In the literature these basic domains are considered to be of primary importance. In Miller & Love (1967) only the irreps for the $\mathbf{k}$ vectors of the basic domains Ω are listed. Section 5.5 of BC and Davies & Cracknell (1976) state that such a listing is not sufficient for the non-holosymmetric space groups because $\Omega < \Phi$. Section 5.5 of BC shows how to overcome this deficiency; Chapter 4 of CDML introduces new labels of $\mathbf{k}$ vectors for the parts of Φ not belonging to Ω.

The crystallographic analogue of the representation domain in direct space is the *asymmetric unit*, *cf. IT* A. According to its definition it is a simply connected smallest part of space from which by application of all symmetry operations of the space group the whole space is exactly filled. For each space-group type the asymmetric units of *IT* A belong to the same topological type independent of the lattice parameters. They are chosen as 'simple' bodies by inspection rather than by applying clearly stated rules. Among the asymmetric units of the 73 symmorphic space-group types $\mathcal{G}_0$ there are 31 parallelepipeds, 27 prisms (13 trigonal, 6 tetragonal and 8 pentagonal) for the non-cubic, and 15 pyramids (11 trigonal and 4 tetragonal) for the cubic $\mathcal{G}_0$.

The asymmetric units of *IT* A – transferred to the groups $(\mathcal{G})^*$ of reciprocal space – are alternatives for the representation domains of the literature. They are formulated as closed bodies. Therefore, for inner points $\mathbf{k}$, the asymmetric units of *IT* A fulfil the condition that each star of $\mathbf{k}$ is represented exactly once. For the surface, however, these conditions either have to be worked out or one gives up the condition of uniqueness and replaces *exactly* by *at least* in the definition of the minimal domain (see preceding footnote). The examples of Section 1.5.5 show that the conditions for the boundary of the asymmetric unit and its special points, lines and planes are in many cases much easier to formulate than those for the representation domain. The asymmetric units used here are not always those of *IT* A. They may have been chosen differently in order to make them more convenient for our purpose or to adapt them more or less to the representation domains of CDML.

The $\mathbf{k}$*-vector coefficients.* For each $\mathbf{k}$ vector one can derive a set of irreps of the space group $\mathcal{G}$. Different $\mathbf{k}$ vectors of a $\mathbf{k}$ orbit give rise to equivalent irreps. Thus, for the calculation of the irreps of the space groups it is essential to identify the orbits of $\mathbf{k}$ vectors in reciprocal space. This means finding the sets of all $\mathbf{k}$ vectors that are related by the operations of the reciprocal-space group $(\mathcal{G})^*$ according to equation (1.5.3.13). The classification of these $\mathbf{k}$ orbits can be done in analogy to that of the point orbits of the symmorphic space groups, as is apparent from the comparison of equations (1.5.3.14) and (1.5.3.15).

The classes of point orbits in direct space under a space group $\mathcal{G}$ are well known and are listed in the space-group tables of *IT* A. They are labelled by *Wyckoff letters*. The stabilizer $\mathcal{S}_{\mathcal{G}}(X)$ of a point X is called the *site-symmetry group* of X, and a *Wyckoff position* consists of all orbits for which the site-symmetry groups are conjugate subgroups of $\mathcal{G}$. Let $\mathcal{G}$ be a symmorphic space group $\mathcal{G}_0$. Owing to the isomorphism between the reciprocal-space groups $(\mathcal{G})^*$ and the symmorphic space groups $\mathcal{G}_0$, the complete list of special $\mathbf{k}$ vectors of $(\mathcal{G})^*$ is provided by the Wyckoff positions of $\mathcal{G}_0$. The groups $\mathcal{S}_{\mathcal{G}_0}(X)$ and $\overline{\mathcal{G}}^{\mathbf{k}}$ correspond to each other and the multiplicity of the Wyckoff position (divided by the number of centring vectors per unit cell for centred lattices) equals the number of arms of the star of $\mathbf{k}$. Let the vectors $\mathbf{t}$ of $\mathbf{L}$ be referred to the conventional basis $(\mathbf{a}_i)^T$ of the space-group tables of *IT* A, as defined in Chapters 2.1 and 9.1 of *IT* A. Then, for the construction of the irreps $\Gamma^{\mathbf{k}}$ of $\mathcal{T}$ the coefficients of the $\mathbf{k}$ vectors must be referred to the basis $(\mathbf{a}_j^*)$ of reciprocal space dual to $(\mathbf{a}_i)^T$ in direct space. These $\mathbf{k}$-vector coefficients may be different from the conventional coordinates of $\mathcal{G}_0$ listed in the Wyckoff positions of *IT* A.

Example

Let $\mathcal{G}$ be a space group with an *I*-centred cubic lattice $\mathbf{L}$, conventional basis $(\mathbf{a}_i)^T$. Then $\mathbf{L}^*$ is an *F*-centred lattice. If referred to the conventional basis $(\mathbf{a}_j^*)$ with $\mathbf{a}_i \cdot \mathbf{a}_j^* = 2\pi\delta_{ij}$, the $\mathbf{k}$ vectors with coefficients 1 0 0, 0 1 0 and 0 0 1 do not belong to $\mathbf{L}^*$ due to the 'extinction laws' well known in X-ray crystallography. However, in the standard basis of $\mathcal{G}_0$, isomorphic to $(\mathcal{G})^*$, the vectors 1 0 0, 0 1 0 and 0 0 1 point to the vertices of the face-centred cube and thus correspond to 2 0 0, 0 2 0 and 0 0 2 referred to the conventional basis $(\mathbf{a}_j^*)$.

In the following, three bases and, therefore, three kinds of coefficients of $\mathbf{k}$ will be distinguished:

(1) Coefficients referred to the conventional basis $(\mathbf{a}_j^*)$ in reciprocal space, dual to the conventional basis $(\mathbf{a}_i)^T$ in direct space. The corresponding $\mathbf{k}$-vector coefficients, $(k_j)^T$, will be called *conventional* coefficients.

(2) Coefficients of $\mathbf{k}$ referred to a primitive basis $(\mathbf{a}_{pi}^*)$ in reciprocal space (which is dual to a primitive basis in direct space). The corresponding coefficients will be called *primitive* coefficients $(k_{pi})^T$. For a centred lattice the coefficients $(k_{pi})^T$ are different from the conventional coefficients $(k_i)^T$. In most of the physics literature related to space-group representations these primitive coefficients are used, *e.g.* by CDML.

(3) The coefficients of $\mathbf{k}$ referred to the conventional basis of $\mathcal{G}_0$. These coefficients will be called *adjusted* coefficients $(k_{ai})^T$.

The relations between conventional and adjusted coefficients are listed for the different Bravais types of reciprocal lattices in Table 1.5.4.1, and those between adjusted and primitive coordinates in Table 1.5.4.2. If adjusted coefficients are used, then *IT* A is as suitable for dealing with irreps as it is for handling space-group symmetry.

1.5.4.3. Wintgen positions

In order to avoid confusion, in the following the analogues to the Wyckoff positions of $\mathcal{G}_0$ will be called *Wintgen positions* of $(\mathcal{G})^*$; the *coordinates* of the Wyckoff position are replaced by the $\mathbf{k}$*-vector coefficients* of the Wintgen position, the *Wyckoff letter* will be called the *Wintgen letter*, and the symbols for the *site symmetries* of $\mathcal{G}_0$ are to be read as the symbols for the *little cogroups* $\overline{\mathcal{G}}^{\mathbf{k}}$ of the $\mathbf{k}$ vectors in $(\mathcal{G})^*$. The *multiplicity* of a Wyckoff position is retained in the Wintgen symbol in order to facilitate the use of *IT* A for the description of symmetry in $\mathbf{k}$ space. However, it is equal to the *multiplicity of the star of* $\mathbf{k}$ only in the case of primitive lattices $\mathbf{L}^*$.

In analogy to a Wyckoff position, a Wintgen position is a set of orbits of $\mathbf{k}$ vectors. Each orbit as well as each star of $\mathbf{k}$ can be represented by any one of its $\mathbf{k}$ vectors. The zero, one or two free parameters in the $\mathbf{k}$-vector coefficients correspond to special $\mathbf{k}$ vectors, *i.e.*, they define *symmetry points*, *symmetry lines* or *symmetry planes*. Three free parameters indicate general $\mathbf{k}$ vectors. The different stars of a Wintgen position are obtained by varying the free parameters.

For convenience, we call the set of all $\mathbf{k}$ vectors of a Wintgen position a $\mathbf{k}$*-vector type*. A $\mathbf{k}$-vector type consists of complete orbits of $\mathbf{k}$ vectors and thus of full stars of $\mathbf{k}$ vectors; the different orbits of a $\mathbf{k}$-vector type are distinguished by the value(s) of the variable parameter(s) of their $\mathbf{k}$ vectors. Correspondingly, the

Table 1.5.4.1. *Conventional coefficients* $(k_i)^T$ *of* **k** *expressed by the adjusted coefficients* (k_{ai}) *of IT A for the different Bravais types of lattices in direct space*

Lattice types	k_1	k_2	k_3
aP, *mP*, *oP*, *tP*, *cP*, *rP*	k_{a1}	k_{a2}	k_{a3}
mA, *oA*	k_{a1}	$2k_{a2}$	$2k_{a3}$
mC, *oC*	$2k_{a1}$	$2k_{a2}$	k_{a3}
oF, *cF*, *oI*, *cI*	$2k_{a1}$	$2k_{a2}$	$2k_{a3}$
tI	$k_{a1}+k_{a2}$	$-k_{a1}+k_{a2}$	$2k_{a3}$
hP	$k_{a1}-k_{a2}$	k_{a2}	k_{a3}
hR (hexagonal)	$2k_{a1}-k_{a2}$	$-k_{a1}+2k_{a2}$	$3k_{a3}$

Table 1.5.4.2. *Primitive coefficients* $(k_{pi})^T$ *of* **k** *from CDML expressed by the adjusted coefficients* (k_{ai}) *of IT A for the different Bravais types of lattices in direct space*

Lattice types	k_{p1}	k_{p2}	k_{p3}
aP, *mP*, *oP*, *tP*, *cP*, *rP*	k_{a1}	k_{a2}	k_{a3}
mA, *oA*	k_{a1}	$k_{a2}-k_{a3}$	$k_{a2}+k_{a3}$
mC, *oC*	$k_{a1}-k_{a2}$	$k_{a1}+k_{a2}$	k_{a3}
oF, *cF*	$k_{a2}+k_{a3}$	$k_{a1}+k_{a3}$	$k_{a1}+k_{a2}$
oI, *cI*	$-k_{a1}+k_{a2}+k_{a3}$	$k_{a1}-k_{a2}+k_{a3}$	$k_{a1}+k_{a2}-k_{a3}$
tI	$-k_{a1}+k_{a3}$	$k_{a1}+k_{a3}$	$k_{a2}-k_{a3}$
hP	$k_{a1}-k_{a2}$	k_{a2}	k_{a3}
hR (hexagonal)	$k_{a1}+k_{a3}$	$-k_{a1}+k_{a2}+k_{a3}$	$-k_{a2}+k_{a3}$

irreps belonging to a **k** vector are specified by these parameters and the irreps belonging to a **k**-vector type form a *type of irreps*, Boyle (1986).

It may be advantageous to describe the different stars belonging to a Wintgen position in a *uniform* way. For this purpose one can define:

Definition. Two **k** vectors of a Wintgen position are *uni-arm* if one can be obtained from the other by parameter variation. The *description of the stars* of a Wintgen position is *uni-arm* if the **k** vectors representing these stars are uni-arm.

The uni-arm description is particularly useful to check whether different sets of **k** vectors belong to the same **k**-vector type or not. Because of the shape of the representation domain or of the asymmetric unit, a **k**-vector type may be split into different parts which belong to different arms of different **k**-vector stars. A uni-arm description may be obtained by the introduction of flagpoles and wings, see Section 1.5.5.1.

For the uni-arm description of a Wintgen position it is easy to check whether the parameter ranges for the general or special constituents of the representation domain or asymmetric unit have been stated correctly. For this purpose one may define the *field of* **k** as the parameter space (point, line, plane or space) of a Wintgen position. For the check one determines that part of the field of **k** which is in the unit cell. The order of the little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$ ($\overline{\mathcal{G}}^{\mathbf{k}}$ represents those operations which leave the field of **k** *fixed pointwise*) is divided by the order of the stabilizer of the field in $(\mathcal{G})^*$ [which is the set of all symmetry operations mod (integer translations) which leave the field *invariant as a whole*]. The result gives the independent fraction of the volume of the unit cell or the area of the plane or length of the line.

If the description is not uni-arm, the uni-arm parameter range will be split into the parameter ranges of the different arms. These parameter ranges of the different arms are not necessarily equal, for examples see Section 1.5.5.

Remark. One should avoid the term *equivalent* for the relation between **k** vectors of the same type but with different parameters, as used by Stokes *et al.* (1993) for Λ and *F* or *B*, *C* and *J* of $m\bar{3}mI$, see examples (1) and (2) in Section 1.5.5.4.4. To belong to the same **k**-vector type is only a necessary, not a sufficient, condition for **k**-vector equivalence. On p. 95 of BC is the following definition: 'Two **k** vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ are equivalent if $\mathbf{k}_1 = \mathbf{k}_2 + \mathbf{K}$, where $\mathbf{K} \in \mathbf{L}^*$'. One can express this by saying: 'Two **k** vectors are equivalent if their difference is a vector **K** of the (reciprocal) lattice'. We prefer to extend this equivalence by saying: 'Two **k** vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ are equivalent if and only if they belong to the same orbit of **k**', *i.e* if there is a matrix part $\boldsymbol{W}$ and a vector $\mathbf{K} \in \mathbf{L}^*$ belonging to $\mathcal{G}$ such that $\mathbf{k}_2 = \mathbf{W}\mathbf{k}_1 + \mathbf{K}$, see equation (1.5.3.13). The **k** vectors of Stokes *et al.* (1993) are not equivalent under this definition, see Davies & Dirl (1987). If the representatives of the **k**-vector stars are chosen uni-arm, their non-equivalence is evident.

It must be mentioned that two **k** vectors of the same type are only called equivalent here if they belong to the same *orbit* of **k** vectors, *i.e.* a **k**-vector type is *not* an equivalence class with respect to the definition of equivalence stated in this remark.

There are two main reasons why **k** vectors of the same type split and then have different labels in CDML:

(1) The higher the symmetry of the point group $\overline{\mathcal{G}}$ of $\mathcal{G}$, the higher is the symmetry of the lattice $\mathbf{L}(\mathcal{G})$ and of the reciprocal lattice $\mathbf{L}^*$, and thus of the Brillouin zone of $\mathcal{G}$. As the symmetry of the Brillouin zone increases, the choice of the boundaries of the representation domain and any other minimal domain becomes more and more restricted. This is because a symmetry element (rotation or rotoinversion axis, plane of reflection, centre of inversion) cannot occur in the interior of the minimal domain but only on its boundary. For the arithmetic crystal class $m\bar{3}mI$, for example, all boundary planes, lines and points are fixed, such that all possible minimal domains are equivalent.

For lower point-group symmetries of the fundamental regions, the choice of the minimal domain is less restricted but the Brillouin zones may become more complicated and may even belong to different topological types depending on the ratios of the lattice parameters. Faces and lines on the surface of the Brillouin zone may appear or disappear or change their relative sizes depending on the lattice parameters, causing different descriptions of Wintgen positions.

This does not happen in unit cells or their asymmetric units. Therefore, as already mentioned in Section 1.5.4.1, BC and CDML preferred to replace the different complicated bodies of the Brillouin zones for the possible values of the lattice parameters of the triclinic and monoclinic lattices by the simple primitive unit cells of the reciprocal lattice and to choose the representation domain correspondingly.

(2) For non-holosymmetric space groups the representation domain Φ is a multiple of the basic domain Ω. CDML introduce new letters for **k** vectors which do not belong to Ω. The more symmetry a space group has lost compared to its holosymmetric space group, the more letters of **k** vectors are introduced. The symbols of the basic domain Ω are kept in the names referred to the representation domain Φ. In most cases one can make the new **k** vectors uni-arm to **k** vectors of the basic domain Ω by an appropriate choice of Ω and Φ. Then these **k** vectors belong to the same **k**-vector type and the additional labels can be avoided by *extension of the parameter range* in the **k**-vector space (Boyle, 1986).

Examples where new letters can be avoided by the extension of the parameter range are common, see, *e.g.*, the examples of Section 1.5.5.

In the following example, the introduction of a new name in the transition from a holosymmetric space group to a non-holosymmetric one cannot be avoided because the Wintgen position splits into two positions. We consider the **k** label $Z, \alpha, \frac{1}{2}, 0$ of CDML for the arithmetic crystal class $m\bar{3}mP$, reciprocal-space group $(\mathcal{G})^* = (Pm\bar{3}m)^*$, isomorphic to $Pm\bar{3}m$, Wyckoff position 12 *h* *mm*2 $x, \frac{1}{2}, 0$. In the subgroup $(Pm\bar{3})^*$, this Wintgen position splits into the two positions $\alpha, \frac{1}{2}, 0$ and $\frac{1}{2}, \alpha, 0$, called *Z* and *ZA* by CDML. In the description of *IT* A they are 6 *g* *mm*2 $x, \frac{1}{2}, 0$ and 6 *f* *mm*2 $\frac{1}{2}, x, 0$. In $Pm\bar{3}$, they form two different Wyckoff positions and thus need two different names, as do their Wintgen positions.

1.5.5. Examples and discussion

The comparison of the crystallographic classification scheme with the traditional one is illustrated by four examples from the Bilbao Crystallographic Server (1998). The examples are designated by the arithmetic crystal classes.

To each arithmetic crystal class of space groups, *cf.* Section 1.5.3.2, there belongs exactly one reciprocal-space group $(\mathcal{G})^*$ which is isomorphic to a type of symmorphic space groups $\mathcal{G}_0$, *cf.* Sections 1.5.3.2 and 1.5.3.4.

(1) **k**-vector types of the arithmetic crystal class $m\bar{3}mI$ (space groups $Im\bar{3}m$ and $Ia\bar{3}d$), reciprocal-space group $(\mathcal{G})^*$ isomorphic to $Fm\bar{3}m$. The representation domain $\Phi = \Omega$ is equal to the asymmetric unit, see Fig. 1.5.5.1 and Table 1.5.5.1.

(2) **k**-vector types of the arithmetic crystal class $m\bar{3}I$ (space groups $Im\bar{3}$ and $Ia\bar{3}$), reciprocal-space group $(\mathcal{G})^*$ isomorphic to $Fm\bar{3}$. The representation domain $\Phi > \Omega$ is equal to the asymmetric unit; see Fig. 1.5.5.2 and Table 1.5.5.2.

(3) **k**-vector types of the arithmetic crystal class $4/mmmI$ ($I4/mmm$, $I4/mcm$, $I4_1/amd$ and $I4_1/acd$), reciprocal-space group $(\mathcal{G})^*$ isomorphic to $I4/mmm$. The representation domains $\Phi = \Omega$ are topologically different for different ratios of the lattice parameters a and c whereas the asymmetric units are affinely equivalent; see Figs. 1.5.5.3 and 1.5.5.4 and Tables 1.5.5.3 and 1.5.5.4.

(4) **k**-vector types of the arithmetic crystal class $mm2F$ ($Fmm2$ and $Fdd2$), reciprocal-space group $(\mathcal{G})^*$ isomorphic to $Imm2$. The representation domains $\Phi > \Omega$ are topologically different for different ratios of the lattice parameters a, b and c whereas the asymmetric units are affinely equivalent; see Figs. 1.5.5.5, 1.5.5.6 and 1.5.5.7, and Tables 1.5.5.5, 1.5.5.6 and 1.5.5.7.

These examples consist essentially of figures and tables. The Brillouin zones with the representation domains of CDML together with the asymmetric units are displayed in the figures. In the synoptic tables the correlation between the **k**-vector tables of CDML and the tables of (Wyckoff) positions in *IT* A is presented. One can thus compare the different descriptions and recognize the relations between them. In addition, the parameter ranges of the **k**-vector types in the asymmetric unit are stated. If a **k**-vector type is listed in the table more than once, then the equivalence relations between the **k** vectors are added such that exactly one representative may be selected for each **k**-vector orbit.

1.5.5.1. Guide to the figures

Each figure caption gives the name of the arithmetic crystal class of space groups to which the Brillouin zone belongs. If there is more than one figure for this arithmetic crystal class, then these figures refer to different geometric conditions for the lattice. Therefore, for each of the figures the arithmetic crystal class is followed by the specific conditions for the lattice parameters of this figure, *e.g.* '$c/a < 1$' for Fig. 1.5.5.3 or '$a^{-2} < b^{-2} + c^{-2}$, $b^{-2} < c^{-2} + a^{-2}$ and $c^{-2} < a^{-2} + b^{-2}$' for Fig. 1.5.5.5.

Then the space groups of the arithmetic crystal class are listed with their Hermann–Mauguin symbols, their Schoenflies symbols and their space-group numbers in *IT* A in parentheses. Following this the type of the reciprocal-space group is denoted, *e.g.* '$(Imm2)^*$, No. 44' for the arithmetic crystal class $mm2F$ in Fig. 1.5.5.5, together with the conditions for the lattice parameters of the reciprocal lattice, if any, and the number of the corresponding table.

The Brillouin zones are objects in reciprocal space. They are displayed in the figures. The reciprocal space is a vector space and its elements are the **k** vectors. Thus the Brillouin zone is a construction in vector space. Because the Brillouin zones are visualized by drawings consisting of vertices, lines and planes, one usually speaks of points, lines and planes in or on the border of the Brillouin zone, not of vectors. Here we follow this tradition.

The Brillouin zones are projected onto the drawing plane by a clinographic projection which may be found *e.g.* in Smith (1982), pp. 61 *f*. The coordinate axes are designated k_x, k_y and k_z; the k_z-coordinate axis points upwards in the projection plane. The diagrams of the Brillouin zones follow those of CDML in order to facilitate the comparison of the data. The origin O with coordinates 0, 0, 0 always forms the centre of the Brillouin zone and is called Γ.

A minimal domain is the smallest fraction of the Brillouin zone which contains *exactly* one wavevector **k** from each orbit. In these examples, the representation domain of CDML is compared with the minimal domain, called 'asymmetric unit', of the Bilbao Crystallographic Server. This asymmetric unit is a simple body and is often chosen in analogy to that of *IT* A. It may coincide with the representation domain of Table 3.10 in CDML, but is mostly rather different. Other than the representation domains of CDML, the asymmetric unit is often *not* fully contained in the Brillouin zone but protrudes from it, in particular by flagpoles and wings, *cf.* the end of this section.

In the figures the edges of the chosen asymmetric unit are drawn into the frame of the Brillouin zone. The names of points, lines and planes of CDML are retained in this listing. New names have been given to points and lines which are not listed in CDML.

The shape of the Brillouin zone depends on the lattice relations. Therefore, there may be vertices of the Brillouin zone with a variable coordinate. If such a point is displayed and designated in a figure by an upper-case letter, then the label of its variable coordinate in the corresponding table is the same letter but lower case. Thus, the variable coordinate of the point G_0 is g_0, of Λ_0 is λ_0 *etc*.

In CDML, the same letter may designate items of different quality in different figures and tables. For example, there is a point H in Fig. 1.5.5.1 and Table 1.5.5.1 but a line H in Fig. 1.5.5.5 and Table 1.5.5.5. In the figures and tables of these examples not only lines and points but also their equivalent objects are listed and the parameter ranges of the lines are described. Therefore, the endpoints of the line H, the points equivalent to a point H as well as the lines equivalent to a line H may be also designated by the letter H but distinguished by indices. In order to recognize points and lines easily, the indices of points are always even: H_0, H_2, H_4; those of lines are always odd: H_1, H_3.

A point is marked in a figure by its name and by a black circle filled with white if it is listed in the corresponding **k**-vector table but is not a point of special symmetry. The same designation is used for the auxiliary points that have been added in order to facilitate the comparison between the two descriptions of the **k**-vector types. Non-coloured parts of the coordinate axes, of the edges of the Brillouin zone or auxiliary lines are displayed by thin solid black lines. Such lines are dashed or omitted if they are not visible, *i.e.* are hidden by the body of the Brillouin zone or of the asymmetric unit.

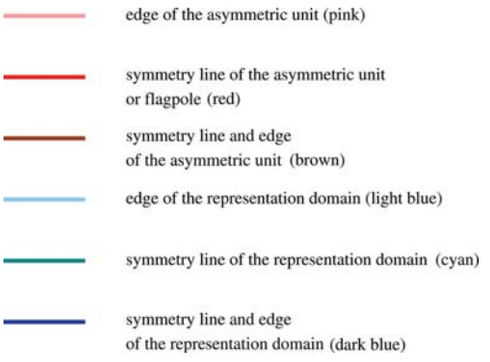
The representatives for the orbits of symmetry points or of symmetry lines, as well as the edges of the representation domain of CDML and of the chosen asymmetric unit are shown in colour.

(*a*) A representative point of each orbit of symmetry points is designated by a red- or cyan-filled circle with its name also in red or cyan if it belongs to the asymmetric unit or to the representation domain of CDML. If both colours could be used, *e.g.* if the asymmetric unit coincides with the representation domain, the colour is red.

Note that a point is coloured red or cyan only if it is really a symmetry *point*, *i.e.* its little co-group is a proper supergroup of the little co-groups of all points in its neighbourhood. Such a point has no variable parameters in its coordinates. Points listed by CDML are not coloured if they are part of a symmetry line or symmetry plane only.

(*b*) Coloured lines are always broad. They are solid lines if they are 'visible', *i.e.* if they are not hidden by the Brillouin zone or by the asymmetric unit. A hidden symmetry line or edge of the asymmetric unit is not suppressed but is coloured as a *dashed line*.

(*c*) The meanings of the different coloured lines and the names used for them in the text are as follows:


edge of the asymmetric unit (pink)
symmetry line of the asymmetric unit or flagpole (red)
symmetry line and edge of the asymmetric unit (brown)
edge of the representation domain (light blue)
symmetry line of the representation domain (cyan)
symmetry line and edge of the representation domain (dark blue)

Notes:

(1) The colour of the line is pink for an edge of the asymmetric unit which is not a symmetry line.

(2) The colour is red for a symmetry line of the asymmetric unit, with the name also in red.

(3) The colour of the line is brown with the name in red for a line which is a symmetry line as well as an edge of the asymmetric unit.

The representation domain of CDML is displayed in the same figure.

(1) The edges of the representation domain are coloured light blue.

(2) The symmetry points and lines with their letters are coloured cyan.

(3) Edges of the representation domain or common edges of the representation domain and the asymmetric unit are coloured dark blue with the letters in cyan if they are symmetry lines of the representation domain but not of the asymmetric unit.

Common edges of an asymmetric unit and a representation domain are coloured pink if they are not symmetry lines simultaneously.

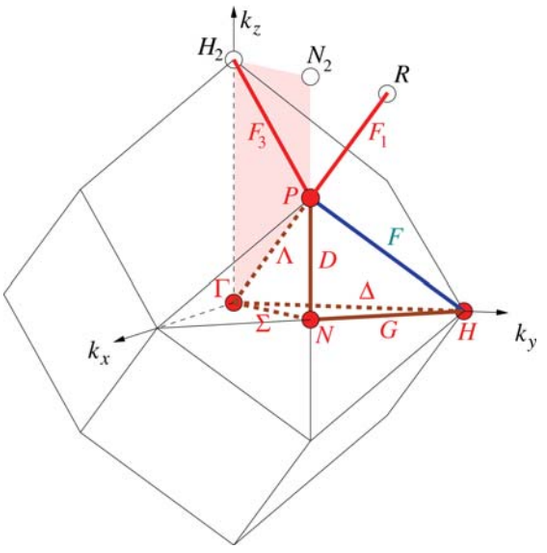


Fig. 1.5.5.1. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class $m\bar{3}mI$. Space groups: $Im\bar{3}m - O_h^9$ (229), $Ia\bar{3}d - O_h^{10}$ (230). Reciprocal-space group $(Fm\bar{3}m)$*, No. 225 (see Table 1.5.5.1). The representation domain of CDML is identical with the asymmetric unit. Auxiliary points: R: $\frac{1}{2},\frac{1}{2},\frac{1}{2}$; N_2: $\frac{1}{4},\frac{1}{4},\frac{1}{2}$; H_2: $0,0,\frac{1}{2}$. Flagpole: $F_1 = [P\,R]$ $\quad x,x,x$: $\frac{1}{4} < x < \frac{1}{2}$. Wing: $B_1 \cup J_1 = [\Gamma\,P\,N_2\,H_2]\backslash[F_3]$ $\quad x,x,z$: $0 < x < \frac{1}{4}$; $x < z < \frac{1}{2}$ with $z \neq \frac{1}{2} - x$.

Exactly one element of each point orbit, line orbit or orbit of planes is contained in the asymmetric unit. Exceptionally, *different* elements of the *same* orbit have been coloured because of their special meaning. In these cases the different elements are connected in the corresponding table by the equivalence sign $\sim$, see, *e.g.* the lines $F \sim F_1 = [P\,R]$ or the planes $B \sim B_1 = [P\,N_2\,H_2]$ in Table 1.5.5.1.

To enable a uni-arm description, symmetry lines outside the asymmetric unit may be selected as orbit representatives. Such a piece of a line is called a *flagpole*. Flagpoles are always coloured red, see, *e.g.*, the line F_1 in Fig. 1.5.5.1.

Symmetry planes are not distinguished in the figures. However, in analogy to the flagpoles, symmetry planes outside the asym-

Table 1.5.5.1. *List of* **k***-vector types for arithmetic crystal class* $m\bar{3}mI$

See Fig. 1.5.5.1. Parameter relations: $x = \frac{1}{2}\beta + \frac{1}{2}\gamma$, $y = \frac{1}{2}\alpha + \frac{1}{2}\gamma$, $z = \frac{1}{2}\alpha + \frac{1}{2}\beta$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3	Parameters
Γ $0,0,0$	4 *a* $m\bar{3}m$	$0,0,0$
H $\frac{1}{2},-\frac{1}{2},\frac{1}{2}$	4 *b* $m\bar{3}m$	$0,\frac{1}{2},0$
P $\frac{1}{4},\frac{1}{4},\frac{1}{4}$	8 *c* $\bar{4}3m$	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$
N $0,0,\frac{1}{2}$	24 *d* $m.mm$	$\frac{1}{4},\frac{1}{4},0$
Δ $\alpha,-\alpha,\alpha$	24 *e* $4m.m$	$0,y,0$: $0 < y < \frac{1}{2}$
Λ α,α,α	*ex* 32 *f* $.3m$	x,x,x: $0 < x < \frac{1}{4}$
F $\frac{1}{2}-\alpha,-\frac{1}{2}+3\alpha,\frac{1}{2}-\alpha$	*ex* 32 *f* $.3m$	$x,\frac{1}{2}-x,x$: $0 < x < \frac{1}{4}$
$F \sim F_1 = [P\,R]$		x,x,x: $\frac{1}{4} < x < \frac{1}{2}$
$F \sim F_3 = [P\,H_2]$		$x,x,\frac{1}{2}-x$: $0 < x < \frac{1}{4}$
$\Lambda \cup F_1 = [\Gamma R]\backslash[P]$	32 *f* $.3m$	x,x,x: $0 < x < \frac{1}{2}, x \neq \frac{1}{4}$
D $\alpha,\alpha,\frac{1}{2}-\alpha$	48 *g* $2.mm$	$\frac{1}{4},\frac{1}{4},z$: $0 < z < \frac{1}{4}$
Σ $0,0,\alpha$	48 *h* $m.m2$	$x,x,0$: $0 < x < \frac{1}{4}$
G $\frac{1}{2}-\alpha,-\frac{1}{2}+\alpha,\frac{1}{2}$	48 *i* $m.m2$	$x,\frac{1}{2}-x,0$: $0 < x < \frac{1}{4}$
A $\alpha,-\alpha,\beta$	96 *j* $m..$	$x,y,0$: $0 < x < y < \frac{1}{2}-x$
B $\alpha+\beta,-\alpha+\beta,\frac{1}{2}-\beta$	*ex* 96 *k* $..m$	$x,\frac{1}{2}-x,z$: $0 < z < x < \frac{1}{4}$
$B \sim B_1 = [P\,N_2\,H_2]$		x,x,z: $0 < x < \frac{1}{2}-x < z < \frac{1}{2}$
C α,α,β	*ex* 96 *k* $..m$	x,x,z: $0 < z < x < \frac{1}{4}$
J α,β,α	*ex* 96 *k* $..m$	x,y,x: $0 < x < y < \frac{1}{2}-x$
$J \sim J_1 = [\Gamma\,P\,H_2]$		x,x,z: $0 < x < z < \frac{1}{2}-x$
$C \cup B_1 \cup J_1 = [\Gamma\,N\,N_2\,H_2]\backslash[\Lambda,\ F_3]$	96 *k* $..m$	x,x,z: $0 < x < \frac{1}{4}, 0 < z < \frac{1}{2}$ with $z \neq x, z \neq \frac{1}{2}-x$
GP α,β,γ	192 *l* 1	x,y,z: $0 < z < x < y < \frac{1}{2}-x$

metric unit may be selected as orbit representatives. Such a piece of a plane is called a *wing*. Wings are always coloured pink, see, *e.g.*, Fig. 1.5.5.1.

Within the caption of each figure the following data are listed:

(i) a statement of whether the representation domain and the asymmetric unit are identical or not;

(ii) the coordinates of auxiliary points if not specified in the corresponding table;

(iii) the parameter descriptions of the flagpoles and the wings.

1.5.5.2. Guide to the **k***-vector tables*

Each figure is followed by a table with the same number. As for the figures, each table caption gives the name of the arithmetic crystal class of space groups. If there is more than one table for this arithmetic crystal class, then the symbol for the arithmetic crystal class is followed by the specific conditions for the lattice parameters, as for the figures.

Column 1. Label of the **k** vectors in CDML, Tables 3.9 and 3.11 and parameter description of CDML for the set of **k** vectors which belong to the label. No ranges for the parameters are listed in CDML.

If two **k** vectors belong to the same type of **k** vectors, then their little co-groups are conjugate under the reciprocal-space group $(\mathcal{G})^*$ and they correspond to the same Wyckoff position. Different **k** vectors with the *same* CDML label always belong to the same **k**-vector type. **k** vectors with *different* CDML labels may either belong to the same or to different types of **k** vectors. If such **k** vectors belong to the same type, the corresponding Wyckoff-position descriptions are preceded by the letters '*ex*'. Frequently, such **k** vectors have been transformed (sign '$\sim$' in these tables) to equivalent ones in order to make the **k** vectors uni-arm, see the tables in this section.

The parameter range of a region may be described by the vertices of that region in brackets [...]. One point in brackets, *e.g.* [*P*], means the point *P*. Two points within the brackets, *e.g.* [*A B*] means the line from *A* to *B*. Three points within the brackets, *e.g.* [*A B C*] means the triangular region of a plane with the vertices *A*, *B* and *C*. Four or more points may mean a region of a plane or a three-dimensional body, depending on the positions of the points. The meaning can be recognized by studying the corresponding figure. Commas between the points, *e.g.* [*A*, *B*, *C*] indicate the set {*A*, *B*, *C*} of the three points *A*, *B* and *C*.

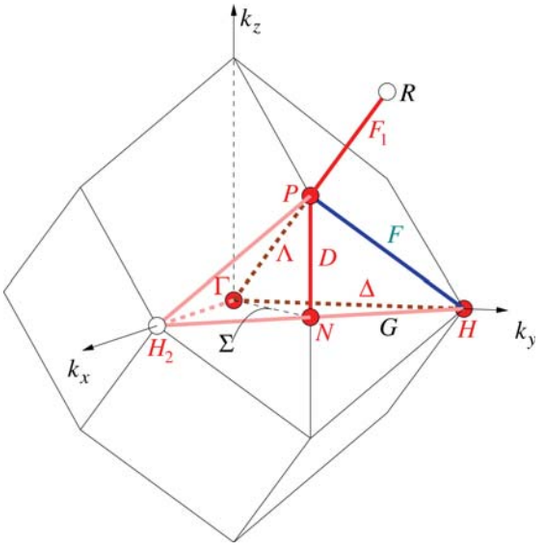


Fig. 1.5.5.2. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class $m\bar{3}I$. Space groups $Im\bar{3}-T_h^5$ (204), $Ia\bar{3}-T_h^7$ (206). Reciprocal-space group $(Fm\bar{3})^*$, No. 202 (see Table 1.5.5.2). The representation domain of CDML is identical with the asymmetric unit. Auxiliary points: R: $\frac{1}{2},\frac{1}{2},\frac{1}{2}$; H_2: $\frac{1}{2},0,0$. Flagpole: $F_1=[P\,R]$ x,x,x: $\frac{1}{4}<x<\frac{1}{2}$.

A symbol [...] does not indicate whether the vertices, boundary lines or boundary planes of the region are themselves included or not. All or part of them may belong to the region, all or part of them may not. In the parameter description of the region in Column 3 the inclusion or exclusion is stated by the symbols $\leq$ or $<$.

The backslash '\' is used to indicate included parts not belonging to the described region, see *e.g.* the regions $[\Gamma\,R]\backslash[P]$ and $[\Gamma\,N\,N_2\,H_2]\backslash[\Lambda,\ F_3]$ in Table 1.5.5.1.

Column 2. This column describes the Wyckoff positions (given as the multiplicity, the Wyckoff letter and the site symmetry) of that symmorphic space group $\mathcal{G}_0$ of *IT* A which is isomorphic to the reciprocal-space group $(\mathcal{G})^*$. Each Wyckoff position of $\mathcal{G}_0$ corresponds to a Wintgen position of $(\mathcal{G})^*$, *i.e.* to a type of **k** vectors of $(\mathcal{G})^*$ and *vice versa*.

'Multiplicity' is the number of points in the conventional unit cell of *IT* A. Here it is the number of arms of the star of the **k**

Table 1.5.5.2. *List of* **k***-vector types for arithmetic crystal class* $m\bar{3}I$

See Fig. 1.5.5.2. Parameter relations: $x=\frac{1}{2}\beta+\frac{1}{2}\gamma$, $y=\frac{1}{2}\alpha+\frac{1}{2}\gamma$, $z=\frac{1}{2}\alpha+\frac{1}{2}\beta$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3	Parameters
Γ 0, 0, 0	4 *a* $m\bar{3}$.	0, 0, 0
H $\frac{1}{2},-\frac{1}{2},\frac{1}{2}$	4 *b* $m\bar{3}$.	$0,\frac{1}{2},0$
P $\frac{1}{4},\frac{1}{4},\frac{1}{4}$	8 *c* 23.	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$
N $0,0,\frac{1}{2}$	24 *d* 2/*m*..	$\frac{1}{4},\frac{1}{4},0$
Δ $\alpha,-\alpha,\alpha$	24 *e* *mm*2..	$0,y,0$: $0<y<\frac{1}{2}$
Λ α,α,α	*ex* 32 *f* .3.	x,x,x: $0<x<\frac{1}{4}$
F $\frac{1}{2}-\alpha,-\frac{1}{2}+3\alpha,\frac{1}{2}-\alpha$	*ex* 32 *f* .3.	$x,\frac{1}{2}-x,x$: $0<x<\frac{1}{4}$
$F\sim F_1=[P\,R]$		x,x,x: $\frac{1}{4}<x<\frac{1}{2}$
$\Lambda\cup F_1\sim[\Gamma\,R]\backslash[P]$	32 *f* .3.	x,x,x: $0<x<\frac{1}{2}$, $x\neq\frac{1}{4}$
D $\alpha,\alpha,\frac{1}{2}-\alpha$	48 *g* 2..	$\frac{1}{4},\frac{1}{4},z$: $0<z<\frac{1}{4}$
Σ $0,0,\alpha$	*ex* 48 *h* *m*..	$x,x,0$: $0<x<\frac{1}{4}$
G $\frac{1}{2}-\alpha,-\frac{1}{2}+\alpha,\frac{1}{2}$	*ex* 48 *h* *m*..	$x,\frac{1}{2}-x,0$: $0<x<\frac{1}{4}$
$A=[\Gamma\,N\,H]$ $\alpha,-\alpha,\beta$	*ex* 48 *h* *m*..	$x,y,0$: $0<x<y<\frac{1}{2}-x$
$AA=[\Gamma\,H_2\,N]$ $-\alpha,\alpha,\beta$	*ex* 48 *h* *m*..	$x,y,0$: $0<y<x<\frac{1}{2}-y$
$\Sigma\cup G\cup A\cup AA$	48 *h* *m*..	$x,y,0$: $0<y<\frac{1}{2}-x<\frac{1}{2}\cup$ $\cup\ x,\frac{1}{2}-x,0$: $0<x<\frac{1}{4}$
GP α,β,γ	96 *i* 1	x,y,z: $0<z\leq x<y<\frac{1}{2}-x\cup$ $\cup\ x,y,z$: $0<z<y<x\leq\frac{1}{2}-y\cup$ $\cup\ x,x,z$: $0<z<x<\frac{1}{4}$

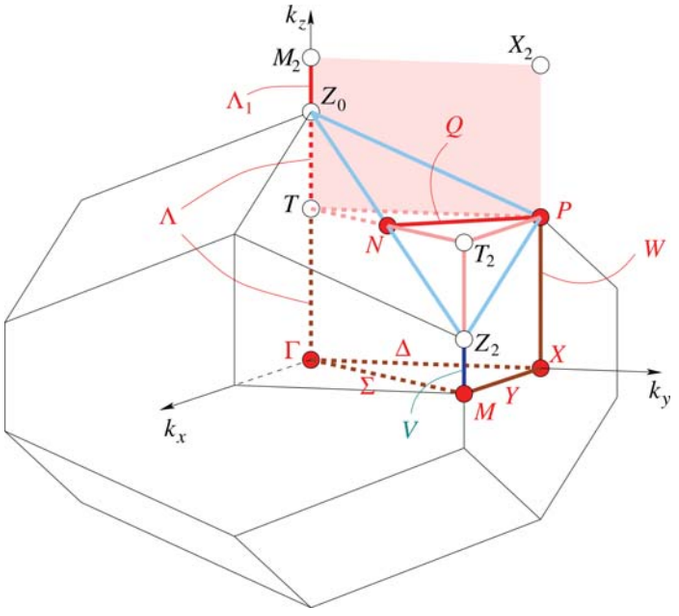


Fig. 1.5.5.3. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class $4/mmmI$: $c/a<1$. Space groups $I4/mmm-D_{4h}^{17}$ (139) to $I4_1/acd-D_{4h}^{20}$ (142). Reciprocal-space group $(I4/mmm)$*, No. 139: $c^*/a^*>1$ (see Table 1.5.5.3). The representation domain of CDML is different from the asymmetric unit. Auxiliary points: T: $0,0,\frac{1}{4}$; T_2: $\frac{1}{2},\frac{1}{2},\frac{1}{4}$; X_2: $0,\frac{1}{2},\frac{1}{2}$. Flagpole: $[T\,M_2]$ $\quad 0,0,z$: $\frac{1}{4}<z<\frac{1}{2}$. Wing: $[T\,P\,X_2\,M_2]$ $\quad 0,y,z$: $0<y<\frac{1}{2}$, $\frac{1}{4}<z<\frac{1}{2}$.

vector, multiplied by the number of centring vectors of the conventional unit cell in *IT* A.

Unlike in *IT* A, each table starts with the Wyckoff letter *a* for a Wyckoff position of highest site symmetry and proceeds in alphabetical order until the general position *GP* is reached. The sequence of the CDML labels is not that of CDML but is determined essentially by the alphabetical sequence of the Wyckoff positions.

The symbol for the site symmetry is 'oriented', as given in the space-group tables of *IT* A. For the nomenclature, see Section 2.2.12 of *IT* A.

Column 3. These are the parameters of that Wyckoff position of $\mathcal{G}_0$ which corresponds to the **k**-vector label in CDML, see Column 1. The *parameter description* and the *parameter range* are listed. This range is chosen such that each orbit of the Wyckoff position of *IT* A, *i.e.* also each **k**-vector orbit, is listed exactly once.

The following designation is used for the parameter ranges:

(1) The statement $0<x,\ y<\frac{1}{2}$ means that x and y may vary independently from 0 to $\frac{1}{2}$, 0 and $\frac{1}{2}$ both excluded.

(2) The statement

$$GP \quad \alpha,\beta,\gamma \quad 48\ h\ 1 \quad x,y,z\colon 0\le z<x<y<\tfrac{1}{2} \ \cup \ \cup\ x,\tfrac{1}{2},z\colon 0<z<x<\tfrac{1}{2}$$

means that the description of the asymmetric unit is split into two adjacent regions, a body and a plane. The boundary plane $z=0$ of the body is included, all other boundaries are excluded. Together the regions contain exactly one representative for each **k**-vector orbit of the general position *GP* of the reciprocal-space group.

(3) The statement $x,\frac{1}{2},z\colon -x<z\le x,\ z\ne 0$ means that z may assume any value between $-x$ and $+x$, $z=x$ included but $z=-x$ and $z=0$ excluded.

(4) Occasionally the parameter description becomes too clumsy. Then the data listed are abbreviated by replacing the parametrical data by the designation of the corresponding region.

Table 1.5.5.3. *List of* ***k****-vector types for arithmetic crystal class* $4/mmmI$: $c/a<1$

See Fig. 1.5.5.3. Wyckoff positions *e* and *f* exchanged. Parameter relations: $x=-\frac{1}{2}\alpha+\frac{1}{2}\beta$, $y=\frac{1}{2}\alpha+\frac{1}{2}\beta+\gamma$, $z=\frac{1}{2}\alpha+\frac{1}{2}\beta$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3				Parameters
$\Gamma\quad 0,0,0$		2	*a*	$4/mmm$	$0,0,0$
$M\quad -\frac{1}{2},\frac{1}{2},\frac{1}{2}$		2	*b*	$4/mmm$	$\frac{1}{2},\frac{1}{2},0$
$M\sim M_2$					$0,0,\frac{1}{2}$
$X\quad 0,0,\frac{1}{2}$		4	*c*	$mmm.$	$0,\frac{1}{2},0$
$P\quad \frac{1}{4},\frac{1}{4},\frac{1}{4}$		4	*d*	$\bar{4}m2$	$0,\frac{1}{2},\frac{1}{4}$
$N\quad 0,\frac{1}{2},0$		8	*f*	$..2/m$	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$
$\Lambda\quad \alpha,\alpha,-\alpha$	*ex*	4	*e*	$4mm$	$0,0,z$: $0<z\le z_0$
$V\quad -\frac{1}{2}+\alpha,\frac{1}{2}+\alpha,\frac{1}{2}-\alpha$	*ex*	4	*e*	$4mm$	$\frac{1}{2},\frac{1}{2},z$: $0<z<z_2=\frac{1}{2}-z_0$
$V\sim\Lambda_1=[Z_0\,M_2]$					$0,0,z$: $z_0<z<\frac{1}{2}$
$\Lambda\cup\Lambda_1=[\Gamma\,M_2]$		4	*e*	$4mm$	$0,0,z$: $0<z<\frac{1}{2}$
$W\quad \alpha,\alpha,\frac{1}{2}-\alpha$		8	*g*	$2mm.$	$0,\frac{1}{2},z$: $0<z<\frac{1}{4}$
$\Sigma\quad -\alpha,\alpha,\alpha$		8	*h*	$m.2m$	$x,x,0$: $0<x<\frac{1}{2}$
$\Delta\quad 0,0,\alpha$		8	*i*	$m2m.$	$0,y,0$: $0<y<\frac{1}{2}$
$Y\quad -\alpha,\alpha,\frac{1}{2}$		8	*j*	$m2m.$	$x,\frac{1}{2},0$: $0<x<\frac{1}{2}$
$Q\quad \frac{1}{4}-\alpha,\frac{1}{4}+\alpha,\frac{1}{4}-\alpha$		16	*k*	$..2$	$x,\frac{1}{2}-x,\frac{1}{4}$: $0<x<\frac{1}{4}$
$C\quad -\alpha,\alpha,\beta$		16	*l*	$m..$	$x,y,0$: $0<x<y<\frac{1}{2}$
$B\quad \alpha,\beta,-\alpha$		16	*m*	$..m$	x,x,z: $[\Gamma\,M\,Z_2\,Z_0]$
$B=B_1\cup B_2$ $=[\Gamma\,M\,Z_2\,N\,T]\cup[T\,N\,Z_0]$					
$B_2\sim B_3$					x,x,z: $[N\,Z_2\,T_2]$
$B_1\cup B_3=[\Gamma\,M\,T_2\,T]$		16	*m*	$..m$	x,x,z: $0<x<\frac{1}{2},\ 0<z<\frac{1}{4}\ \cup$ $\cup\ x,x,\frac{1}{4}$: $0<x<\frac{1}{4}$
$A\quad \alpha,\alpha,\beta$	*ex*	16	*n*	$.m.$	$0,y,z$: $[\Gamma\,X\,P\,Z_0]$
$E\quad \alpha-\beta,\alpha+\beta,\frac{1}{2}-\alpha$	*ex*	16	*n*	$.m.$	$x,\frac{1}{2},z$: $[M\,X\,P\,Z_2]$
$E\sim A_1$					$0,y,z$: $[P\,X_2\,M_2\,Z_0]$
$A\cup A_1=[\Gamma\,X\,X_2\,M_2]$		16	*n*	$.m.$	$0,y,z$: $0<y,z<\frac{1}{2}$
$GP\quad \alpha,\beta,\gamma$		32	*o*	1	x,y,z: $0<x<y<\frac{1}{2}$; $0<z<\frac{1}{4}\ \cup$ $\cup\ x,y,\frac{1}{4}$: $0<x<y<\frac{1}{2}-x$

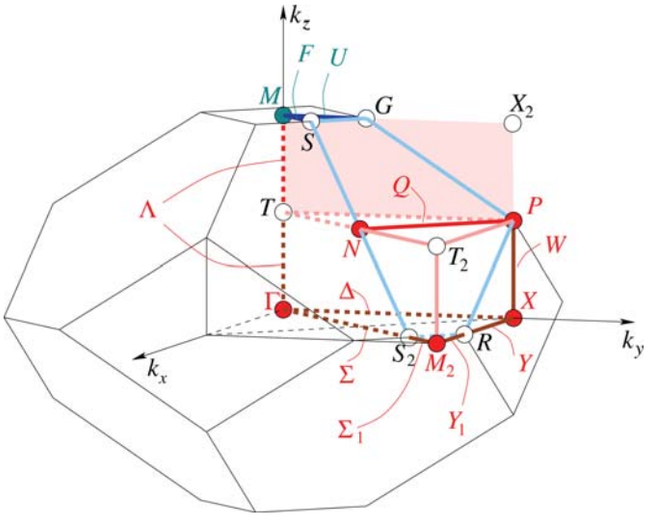


Fig. 1.5.5.4. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class $4/mmmI$: $c/a>1$. Space groups $I4/mmm-D_{4h}^{17}$ (139) to $I4_1/acd-D_{4h}^{20}$ (142). Reciprocal-space group $(I4/mmm)$*, No. 139: $c^*/a^*<1$ (see Table 1.5.5.4). The representation domain of CDML is different from the asymmetric unit. Auxiliary points: X_2: $0,\frac{1}{2},\frac{1}{2}$; T: $0,0,\frac{1}{4}$; T_2: $\frac{1}{2},\frac{1}{2},\frac{1}{4}$. Flagpole: $[T\,M]$ $0,0,z$: $\frac{1}{4}<z<\frac{1}{2}$. Wing: $[T\,P\,X_2\,M]$ $0,y,z$: $0<y<\frac{1}{2}, \frac{1}{4}<z<\frac{1}{2}$.

Example. In Table 1.5.5.3 one finds for the arithmetic crystal class $4/mmmI$ of space groups:

$$B\quad \alpha,\beta,-\alpha\quad 16\ m\ ..m\quad x,x,z\text{: }[\Gamma\,M\,Z_2\,Z_0]$$

The parameter description would be:

$$x,x,z\text{: }0<x<\tfrac{1}{2},\ 0<z\le z_0-2x(2z_0-\tfrac{1}{2})$$

Horizontal lines. The horizontal lines extending across the tables separate blocks with different numbers of free parameters. Decisive for this subdivision is the number of free parameters of the Wyckoff position to which the Wintgen position is assigned, not the number of free parameters of CDML.

Example. Arithmetic crystal class mm2F, see Table 1.5.5.5

The **k**-vector labels 'Γ $0,0,0$' and 'Z $\frac{1}{2},\frac{1}{2},0$' of CDML have no free parameter. However, they correspond to the Wyckoff position '2 *a mm*2 $0,0,z$', which has one free parameter. Therefore, Γ and Z are listed together with 'Λ $\alpha,\alpha,0$' and 'LE $-\alpha,-\alpha,0$' in the block for the symmetry lines, *i.e.* for the **k** vectors with one free parameter: in $(Imm2)$* there is no parameter-free Wintgen position at all. The **k**-vector labels 'Σ $0,\alpha,\alpha$' and 'A $\frac{1}{2},\frac{1}{2}+\alpha,\alpha$' of CDML have one free parameter each. However, they correspond together with other **k**-vector labels to the Wyckoff position '4 *c .m.* $x,0,z$'. Therefore, Σ and A are listed together with 'J $\alpha,\alpha+\beta,\beta$' and '$JA$ $-\alpha,-\alpha+\beta,\beta$' and others in the block for the planes, *i.e.* for the **k** vectors with two free parameters.

Table 1.5.5.4. *List of* **k***-vector types for arithmetic crystal class* $4/mmmI$: $c/a>1$

See Fig. 1.5.5.4. Wyckoff positions e and f exchanged. Parameter relations: $x=-\frac{1}{2}\alpha+\frac{1}{2}\beta$, $y=\frac{1}{2}\alpha+\frac{1}{2}\beta+\gamma$, $z=\frac{1}{2}\alpha+\frac{1}{2}\beta$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3	Parameters
Γ $0,0,0$	2 a $4/mmm$	$0,0,0$
M $\frac{1}{2},\frac{1}{2},-\frac{1}{2}$	2 b $4/mmm$	$0,0,\frac{1}{2}$
$M\sim M_2$		$\frac{1}{2},\frac{1}{2},0$
X $0,0,\frac{1}{2}$	4 c $mmm.$	$0,\frac{1}{2},0$
P $\frac{1}{4},\frac{1}{4},\frac{1}{4}$	4 d $\bar{4}m2$	$0,\frac{1}{2},\frac{1}{4}$
N $0,\frac{1}{2},0$	8 f $..2/m$	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$
Λ $\alpha,\alpha,-\alpha$	4 e $4mm$	$0,0,z$: $0<z<\frac{1}{2}$
W $\alpha,\alpha,\frac{1}{2}-\alpha$	8 g $2mm.$	$0,\frac{1}{2},z$: $0<z<\frac{1}{4}$
Σ $-\alpha,\alpha,\alpha$	*ex* 8 h $m.2m$	$x,x,0$: $0<x\le s_2$
F $\frac{1}{2}-\alpha,\frac{1}{2}+\alpha,-\frac{1}{2}+\alpha$	*ex* 8 h $m.2m$	$x,x,\frac{1}{2}$: $0<x<s=\frac{1}{2}-s_2$
$F\sim\Sigma_1=[S_2\,M_2]$		$x,x,0$: $s_2<x<\frac{1}{2}$
$\Sigma\cup\Sigma_1=[\Gamma M_2]$	8 h $m.2m$	$x,x,0$: $0<x<\frac{1}{2}$
Δ $0,0,\alpha$	8 i $m2m.$	$0,y,0$: $0<y<\frac{1}{2}$
Y $-\alpha,\alpha,\frac{1}{2}$	*ex* 8 j $m2m.$	$x,\frac{1}{2},0$: $0<x\le r$
U $\frac{1}{2},\frac{1}{2},-\frac{1}{2}+\alpha$	*ex* 8 j $m2m.$	$0,y,\frac{1}{2}$: $0<y<g=\frac{1}{2}-r$
$U\sim Y_1=[R\,M_2]$		$x,\frac{1}{2},0$: $r<x<\frac{1}{2}$
$Y\cup Y_1=[X\,M_2]$	8 j $m2m.$	$x,\frac{1}{2},0$: $0<x<\frac{1}{2}$
Q $\frac{1}{4}-\alpha,\frac{1}{4}+\alpha,\frac{1}{4}-\alpha$	16 k $..2$	$x,\frac{1}{2}-x,\frac{1}{4}$: $0<x<\frac{1}{4}$
C $-\alpha,\alpha,\beta$	*ex* 16 l $m..$	$x,y,0$: $[\Gamma\,S_2\,R\,X]$
D $\frac{1}{2}-\alpha,\frac{1}{2}+\alpha,-\frac{1}{2}+\beta$	*ex* 16 l $m..$	$x,y,\frac{1}{2}$: $[M\,S\,G]$
$D\sim C_1$		$x,y,0$: $[M_2\,R\,S_2]$
$C\cup C_1=[\Gamma\,M_2\,X]$	16 l $m..$	$x,y,0$: $0<x<y<\frac{1}{2}$
B $\alpha,\beta,-\alpha$	16 m $..m$	x,x,z: $[\Gamma\,S_2\,S\,M]$
$B=B_1\cup B_2$ $=[\Gamma\,S_2\,N\,T]\cup[T\,N\,S\,M]$		
$B_2\sim B_3$		x,x,z: $[T_2\,N\,S_2\,M_2]$
$B_1\cup B_3=[\Gamma\,M_2\,T_2\,T]$	16 m $..m$	x,x,z: $0<x<\frac{1}{2}, 0<z<\frac{1}{4}\cup$ $\cup\,x,x,\frac{1}{4}$: $0<x<\frac{1}{4}$
A α,α,β	*ex* 16 n $.m.$	$0,y,z$: $[\Gamma\,X\,P\,G\,M]$
E $\alpha-\beta,\alpha+\beta,\frac{1}{2}-\alpha$	*ex* 16 n $.m.$	$x,\frac{1}{2},z$: $[X\,P\,R]$
$E\sim A_1$		$0,y,z$: $[X_2\,G\,P]$
$A\cup A_1=[\Gamma\,X\,X_2\,M]$	16 n $.m.$	$0,y,z$: $0<y,z<\frac{1}{2}$
GP α,β,γ	32 o 1	x,y,z: $0<x<y<\frac{1}{2}$; $0<z<\frac{1}{4}\cup$ $\cup\,x,y,\frac{1}{4}$: $0<x<y<\frac{1}{2}-x$

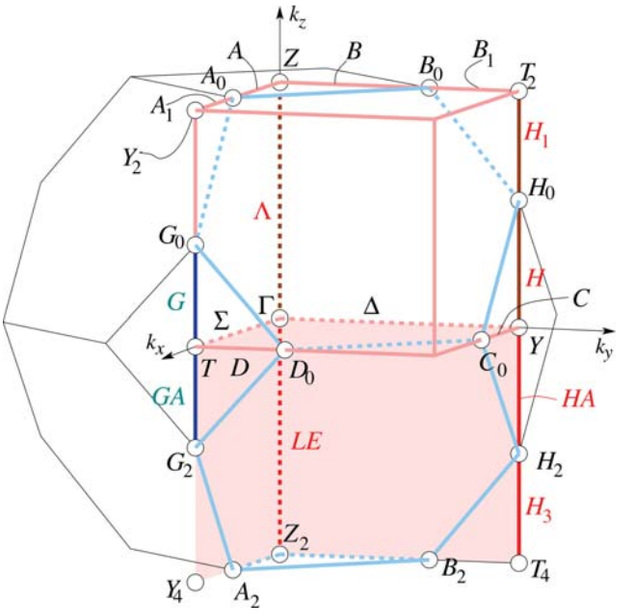


Fig. 1.5.5.5. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class *mm2F*: $a^{-2} < b^{-2} + c^{-2}$, $b^{-2} < c^{-2} + a^{-2}$ and $c^{-2} < a^{-2} + b^{-2}$. Space groups $Fmm2 - C_{2v}^{18}$ (42), $Fdd2 - C_{2v}^{19}$ (43). Reciprocal-space group $(Imm2)^*$, No. 44: $a^{*2} < b^{*2} + c^{*2}$, $b^{*2} < c^{*2} + a^{*2}$ and $c^{*2} < a^{*2} + b^{*2}$ (see Table 1.5.5.5). The representation domain of CDML is different from the asymmetric unit. Auxiliary points: T_4: $0, \frac{1}{2}, -\frac{1}{2}$; Y_2: $\frac{1}{2}, 0, \frac{1}{2}$; Y_4: $\frac{1}{2}, 0, -\frac{1}{2}$; Z_2: $0, 0, -\frac{1}{2}$. Flagpoles: $0, 0, z$: $-\frac{1}{2} < z < 0$; $0, \frac{1}{2}, z$: $-\frac{1}{2} < z < 0$. Wings: $x, 0, z$: $0 < x < \frac{1}{2}, -\frac{1}{2} < z < 0$; $0, y, z$: $0 < y < \frac{1}{2}, -\frac{1}{2} < z < 0$.

Table 1.5.5.5. *List of* ***k****-vector types for arithmetic crystal class mm2F:* $a^{-2} < b^{-2} + c^{-2}$, $b^{-2} < c^{-2} + a^{-2}$ *and* $c^{-2} < a^{-2} + b^{-2}$

See Fig. 1.5.5.5. Parameter relations: $x = -\frac{1}{2}\alpha + \frac{1}{2}\beta + \frac{1}{2}\gamma$, $y = \frac{1}{2}\alpha - \frac{1}{2}\beta + \frac{1}{2}\gamma$, $z = \frac{1}{2}\alpha + \frac{1}{2}\beta - \frac{1}{2}\gamma$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3				Parameters
$\Gamma \quad 0, 0, 0$	*ex*	2	*a*	*mm2*	$0, 0, 0$
$Z \quad \frac{1}{2}, \frac{1}{2}, 0$	*ex*	2	*a*	*mm2*	$0, 0, \frac{1}{2}$
$\Lambda \quad \alpha, \alpha, 0$	*ex*	2	*a*	*mm2*	$0, 0, z$: $0 < z < \frac{1}{2}$
$LE \quad -\alpha, -\alpha, 0$	*ex*	2	*a*	*mm2*	$0, 0, z$: $-\frac{1}{2} < z < 0$
$\Gamma \cup \Lambda \cup Z \cup LE$		2	*a*	*mm2*	$0, 0, z$: $-\frac{1}{2} < z \leq \frac{1}{2}$
$T \quad 0, \frac{1}{2}, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$\frac{1}{2}, 0, 0$
$T \sim T_2$					$0, \frac{1}{2}, \frac{1}{2}$
$Y \quad \frac{1}{2}, 0, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, 0$
$G \quad \alpha, \frac{1}{2} + \alpha, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$\frac{1}{2}, 0, z$: $0 < z \leq g_0$
$G \sim H_3 = [H_2\, T_4]$					$0, \frac{1}{2}, z$: $-\frac{1}{2} < z \leq -\frac{1}{2} + g_0 = h_2$
$GA \quad -\alpha, \frac{1}{2} - \alpha, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$\frac{1}{2}, 0, z$: $g_2 = -g_0 < z < 0$
$GA \sim H_1 = [H_0\, T_2]$					$0, \frac{1}{2}, z$: $\frac{1}{2} - g_0 = h_0 < z < \frac{1}{2}$
$H \quad \frac{1}{2} + \alpha, \alpha, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, z$: $0 < z \leq h_0$
$HA \quad \frac{1}{2} - \alpha, -\alpha, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, z$: $h_2 = -h_0 < z < 0$
$T_2 \cup H_1 \cup H \cup Y \cup HA \cup H_3$		2	*b*	*mm2*	$0, \frac{1}{2}, z$: $-\frac{1}{2} < z \leq \frac{1}{2}$
$\Sigma \quad 0, \alpha, \alpha$	*ex*	4	*c*	*.m.*	$x, 0, 0$: $0 < x < \frac{1}{2}$
$A \quad \frac{1}{2}, \frac{1}{2} + \alpha, \alpha$	*ex*	4	*c*	*.m.*	$x, 0, \frac{1}{2}$: $0 < x \leq a_0$
$C \quad \frac{1}{2}, \alpha, \frac{1}{2} + \alpha$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, 0$: $0 < x < c_0 = \frac{1}{2} - a_0$
$C \sim A_1$					$x, 0, \frac{1}{2}$: $\frac{1}{2} - a_0 = c_0 < x < \frac{1}{2}$
$J \quad \alpha, \alpha + \beta, \beta$	*ex*	4	*c*	*.m.*	$x, 0, z$: $[\Gamma\, Z\, A_0\, G_0\, T]$
$JA \quad -\alpha, -\alpha + \beta, \beta$	*ex*	4	*c*	*.m.*	$x, 0, z$: $[\Gamma\, T\, G_2\, A_2\, Z_2]$
$K \quad \frac{1}{2} + \alpha, \alpha + \beta, \frac{1}{2} + \beta$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, z$: $[Y\, H_0\, C_0]$
$K \sim J_1$					$x, 0, z$: $[Y_4\, G_2\, A_2]$
$KA \quad \frac{1}{2} - \alpha, -\alpha + \beta, \frac{1}{2} + \beta$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, z$: $[Y\, C_0\, H_2]$
$KA \sim J_3$					$x, 0, z$: $[Y_2\, G_0\, A_0]$
$A \cup A_1 \cup J \cup J_3 \cup \Sigma \cup JA \cup J_1$		4	*c*	*.m.*	$x, 0, z$: $0 < x < \frac{1}{2}$; $0 < z \leq \frac{1}{2}$
$\Delta \quad \alpha, 0, \alpha$	*ex*	4	*d*	*m..*	$0, y, 0$: $0 < y < \frac{1}{2}$
$B \quad \frac{1}{2} + \alpha, \frac{1}{2}, \alpha$	*ex*	4	*d*	*m..*	$0, y, \frac{1}{2}$: $0 < y < b_0$
$D \quad \alpha, \frac{1}{2}, \frac{1}{2} + \alpha$	*ex*	4	*d*	*m..*	$\frac{1}{2}, y, 0$: $0 < y \leq d_0$
$D \sim B_1$					$0, y, \frac{1}{2}$: $\frac{1}{2} - d_0 = b_0 \leq y < \frac{1}{2}$
$E \quad \alpha + \beta, \alpha, \beta$	*ex*	4	*d*	*m..*	$0, y, z$: $[\Gamma\, Y\, H_0\, B_0\, Z]$
$EA \quad -\alpha + \beta, -\alpha, \beta$	*ex*	4	*d*	*m..*	$0, y, z$: $[\Gamma\, Z_2\, B_2\, H_2\, Y]$
$F \quad \alpha + \beta, \frac{1}{2} + \alpha, \frac{1}{2} + \beta$	*ex*	4	*d*	*m..*	$\frac{1}{2}, y, z$: $[T\, D_0\, G_0]$
$F \sim E_3$					$0, y, z$: $[B_2\, T_4\, H_2]$
$FA \quad -\alpha + \beta, \frac{1}{2} - \alpha, \frac{1}{2} + \beta$	*ex*	4	*d*	*m..*	$\frac{1}{2}, y, z$: $[T\, G_2\, D_0]$
$FA \sim E_1$					$0, y, z$: $[T_2\, B_0\, H_0]$
$\Delta \cup B \cup B_1 \cup E \cup E_1 \cup EA \cup E_3$		4	*d*	*m..*	$0, y, z$: $0 < y < \frac{1}{2}$; $-\frac{1}{2} < z \leq \frac{1}{2}$
$GP \quad \alpha, \beta, \gamma$		8	*e*	1	x, y, z: $0 < x, y < \frac{1}{2}$; $0 < z \leq \frac{1}{2}$

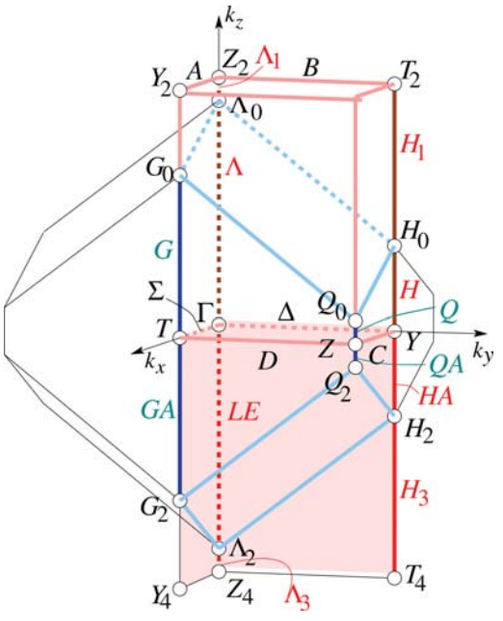


Fig. 1.5.5.6. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class *mm2F*: $c^{-2} > a^{-2} + b^{-2}$. Space groups $Fmm2 - C_{2v}^{18}$ (42), $Fdd2 - C_{2v}^{19}$ (43). Reciprocal-space group (*Imm2*)*, No. 44: $c^{*2} > a^{*2} + b^{*2}$ (see Table 1.5.5.6). The representation domain of CDML is different from the asymmetric unit. Auxiliary points: T_4: $0, \frac{1}{2}, -\frac{1}{2}$; Y_4: $\frac{1}{2}, 0, -\frac{1}{2}$; Z_4: $0, 0, -\frac{1}{2}$. Flagpoles: $0, 0, z$: $-\frac{1}{2} < z < 0$; $0, \frac{1}{2}, z$: $-\frac{1}{2} < z < 0$. Wings: $x, 0, z$: $0 < x < \frac{1}{2}, -\frac{1}{2} < z < 0$; $0, y, z$: $0 < y < \frac{1}{2}, -\frac{1}{2} < z < 0$.

Table 1.5.5.6. *List of* **k***-vector types for arithmetic crystal class mm2F:* $c^{-2} > a^{-2} + b^{-2}$

See Fig. 1.5.5.6. Parameter relations: $x = -\frac{1}{2}\alpha + \frac{1}{2}\beta + \frac{1}{2}\gamma$, $y = \frac{1}{2}\alpha - \frac{1}{2}\beta + \frac{1}{2}\gamma$, $z = \frac{1}{2}\alpha + \frac{1}{2}\beta - \frac{1}{2}\gamma$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3	Parameters
Γ $0, 0, 0$	*ex* 2 *a* *mm2*	$0, 0, 0$
Z $\frac{1}{2}, \frac{1}{2}, 1$	*ex* 2 *a* *mm2*	$\frac{1}{2}, \frac{1}{2}, 0$
$Z \sim Z_2$		$0, 0, \frac{1}{2}$
Λ $\alpha, \alpha, 0$	*ex* 2 *a* *mm2*	$0, 0, z$: $0 < z \le \lambda_0$
LE $-\alpha, -\alpha, 0$	*ex* 2 *a* *mm2*	$0, 0, z$: $\lambda_2 = -\lambda_0 < z < 0$
Q $\frac{1}{2}+\alpha, \frac{1}{2}+\alpha, 1$	*ex* 2 *a* *mm2*	$\frac{1}{2}, \frac{1}{2}, z$: $0 < z \le q_0$
$Q \sim \Lambda_3 = [\Lambda_2\, Z_4]$		$0, 0, z$: $-\frac{1}{2} < z \le -\frac{1}{2} + q_0 = -\lambda_0$
QA $\frac{1}{2}-\alpha, \frac{1}{2}-\alpha, 1$	*ex* 2 *a* *mm2*	$\frac{1}{2}, \frac{1}{2}, z$: $q_2 = -q_0 < z < 0$
$QA \sim \Lambda_1 = [\Lambda_0\, Z_2]$		$0, 0, z$: $\frac{1}{2} - q_0 = \lambda_0 < z < \frac{1}{2}$
$Z_2 \cup \Lambda_1 \cup \Lambda \cup \Gamma \cup LE \cup \Lambda_3$	2 *a* *mm2*	$0, 0, z$: $-\frac{1}{2} < z \le \frac{1}{2}$
T $0, \frac{1}{2}, \frac{1}{2}$	*ex* 2 *b* *mm2*	$\frac{1}{2}, 0, 0$
$T \sim T_2$		$0, \frac{1}{2}, \frac{1}{2}$
Y $\frac{1}{2}, 0, \frac{1}{2}$	*ex* 2 *b* *mm2*	$0, \frac{1}{2}, 0$
G $\alpha, \frac{1}{2}+\alpha, \frac{1}{2}$	*ex* 2 *b* *mm2*	$\frac{1}{2}, 0, z$: $0 < z \le g_0$
$G \sim H_3 = [H_2\, T_4]$		$0, \frac{1}{2}, z$: $-\frac{1}{2} < z \le -\frac{1}{2} + g_0$
GA $-\alpha, \frac{1}{2}-\alpha, \frac{1}{2}$	*ex* 2 *b* *mm2*	$\frac{1}{2}, 0, z$: $g_2 = -g_0 < z < 0$
$GA \sim H_1 = [H_0\, T_2]$		$0, \frac{1}{2}, z$: $\frac{1}{2} - g_0 = h_0 < z < \frac{1}{2}$
H $\frac{1}{2}+\alpha, \alpha, \frac{1}{2}$	*ex* 2 *b* *mm2*	$0, \frac{1}{2}, z$: $0 < z \le h_0$
HA $\frac{1}{2}-\alpha, -\alpha, \frac{1}{2}$	*ex* 2 *b* *mm2*	$0, \frac{1}{2}, z$: $h_2 = -h_0 < z < 0$
$T_2 \cup H_1 \cup H \cup Y \cup HA \cup H_3$	2 *b* *mm2*	$0, \frac{1}{2}, z$: $-\frac{1}{2} < z \le \frac{1}{2}$
Σ $0, \alpha, \alpha$	*ex* 4 *c* *.m.*	$x, 0, 0$: $0 < x < \frac{1}{2}$
C $\frac{1}{2}, \alpha, \frac{1}{2}+\alpha$	*ex* 4 *c* *.m.*	$x, \frac{1}{2}, 0$: $0 < x < \frac{1}{2}$
$C \sim A = [Z_2\, Y_2]$		$x, 0, \frac{1}{2}$: $0 < z < \frac{1}{2}$
J $\alpha, \alpha+\beta, \beta$	*ex* 4 *c* *.m.*	$x, 0, z$: $[\Gamma\, \Lambda_0\, G_0\, T]$
JA $-\alpha, -\alpha+\beta, \beta$	*ex* 4 *c* *.m.*	$x, 0, z$: $[\Gamma\, T\, G_2\, \Lambda_2]$
K $\frac{1}{2}+\alpha, \alpha+\beta, \frac{1}{2}+\beta$	*ex* 4 *c* *.m.*	$x, \frac{1}{2}, z$: $[Y\, H_0\, Q_0\, Z]$
$K \sim J_3$		$x, 0, z$: $[Y_4\, G_2\, \Lambda_2\, Z_4]$
KA $\frac{1}{2}-\alpha, -\alpha+\beta, \frac{1}{2}+\beta$	*ex* 4 *c* *.m.*	$x, \frac{1}{2}, z$: $[Z\, Q_2\, H_2\, Y]$
$KA \sim J_1$		$x, 0, z$: $[Z_2\, Y_2\, G_0\, \Lambda_0]$
$A \cup J_1 \cup J \cup \Sigma \cup JA \cup J_3$	4 *c* *.m.*	$x, 0, z$: $0 < x < \frac{1}{2}$; $-\frac{1}{2} < z \le \frac{1}{2}$
Δ $\alpha, 0, \alpha$	*ex* 4 *d* *m..*	$0, y, 0$: $0 < y < \frac{1}{2}$
D $\alpha, \frac{1}{2}, \frac{1}{2}+\alpha$	*ex* 4 *d* *m..*	$\frac{1}{2}, y, 0$: $0 < y < \frac{1}{2}$
$D \sim B$		$0, y, \frac{1}{2}$: $0 < y < \frac{1}{2}$
E $\alpha+\beta, \alpha, \beta$	*ex* 4 *d* *m..*	$0, y, z$: $[\Gamma\, Y\, H_0\, \Lambda_0]$
EA $-\alpha+\beta, -\alpha, \beta$	*ex* 4 *d* *m..*	$0, y, z$: $[\Gamma\, \Lambda_2\, H_2\, Y]$
F $\alpha+\beta, \frac{1}{2}+\alpha, \frac{1}{2}+\beta$	*ex* 4 *d* *m..*	$\frac{1}{2}, y, z$: $[T\, Z\, Q_0\, G_0]$
$F \sim E_3$		$0, y, z$: $[Z_4\, \Lambda_2\, H_2\, T_4]$
FA $-\alpha+\beta, \frac{1}{2}-\alpha, \frac{1}{2}+\beta$	*ex* 4 *d* *m..*	$\frac{1}{2}, y, z$: $[T\, G_2\, Q_2\, Z]$
$FA \sim E_1$		$0, y, z$: $[Z_2\, \Lambda_0\, H_0\, T_2]$
$B \cup E_1 \cup E \cup \Delta \cup EA \cup E_3$	4 *d* *m..*	$0, y, z$: $0 < y < \frac{1}{2}$; $-\frac{1}{2} < z \le \frac{1}{2}$
GP α, β, γ	8 *e* 1	x, y, z: $0 < x, y < \frac{1}{2}$; $0 < z \le \frac{1}{2}$

In general the sequence of the Wyckoff letters in *IT* A follows the falling number of free parameters. In the few cases where the sequence in *IT* A is different, the Wyckoff letters are exchanged. The exchange is noted at the top of the table.

Example. In the arithmetic crystal class $4/mmmI$, $c/a < 1$, see Table 1.5.5.3, Wyckoff position *e* has one free parameter, whereas Wyckoff position *f* has constant parameters, *i.e.* no free parameter. Therefore, *f* is listed above the horizontal line, *e* is listed below, see Table 1.5.5.3. The note at the top of the table states 'Wyckoff positions *e* and *f* exchanged'.

Parameter relations. The relations between the parameters of CDML and the parameters referred to the asymmetric unit are listed at the top of the table, *e.g.* for $m\bar{3}mI$ in Table 1.5.5.1: 'Parameter relations: $x = \frac{1}{2}\beta + \frac{1}{2}\gamma$, $y = \frac{1}{2}\alpha + \frac{1}{2}\gamma$, $z = \frac{1}{2}\alpha + \frac{1}{2}\beta$'. These relations may be modified to more convenient parameters without notice, as for the plane *B* of $m\bar{3}mI$ in Table 1.5.5.1:

$$B \quad \alpha+\beta, -\alpha+\beta, \tfrac{1}{2}-\beta \quad ex\ 96\ k\ ..m \quad x, \tfrac{1}{2}-x, z\colon\ 0 < z < x < \tfrac{1}{4}$$

instead of

$$\ldots \tfrac{1}{4} - \tfrac{1}{2}\alpha, \tfrac{1}{4} + \tfrac{1}{2}\alpha, \beta\colon\ 0 < \alpha < \tfrac{1}{2} - 2\beta < \tfrac{1}{2}.$$

1.5.5.3. Figures and tables

Arithmetic crystal classes $m\bar{3}mI$ *and* $m\bar{3}I$: The reciprocal lattice of a cubic lattice *I* is a cubic lattice *F*. Its Brillouin zone is a rhombic dodecahedron and has 12 faces, 24 edges and 14 apices, the coordinates of which are the six permutations of $\pm\frac{1}{2}, 0, 0$ and the eight coordinate triplets of $\pm\frac{1}{4}, \pm\frac{1}{4}, \pm\frac{1}{4}$. Eleven of these 14 points are visible in the applied projection.

The figure for arithmetic crystal class $m\bar{3}mI$ is shown in Fig. 1.5.5.1 and the corresponding table is Table 1.5.5.1. The figure for arithmetic crystal class $m\bar{3}I$ is shown in Fig. 1.5.5.2 and the corresponding table is Table 1.5.5.2.

Arithmetic crystal class $4/mmmI$: There are two different types of Brillouin zones for the tetragonal *I* lattice, one for $c < a$ (Fig. 1.5.5.3, Table 1.5.5.3) and one for $c > a$ (Fig. 1.5.5.4, Table 1.5.5.4). The first type of Brillouin zone, Fig. 1.5.5.3, is a tetragonal elongated rhombdodecahedron with 12 faces, four of them being hexagons. There are 18 apices; 14 of them are visible. The Brillouin zone of Fig. 1.5.5.4 is a tetragonally deformed cuboctahedron with 14 faces. There are 24 apices; 18 of them are visible.

Arithmetic crystal class mm2F: Depending on the lattice ratios *a*:*b*:*c*, there are four figures in CDML for the Brillouin zone of an orthorhombic crystal with an *F* lattice, see Fig. 3.6 on p. 26 in CDML. Only three of them are really necessary. Therefore, the case $b^{-2} > c^{-2} + a^{-2}$ of Fig. 3.6(c) of CDML has been omitted in these examples; it is obtained from $a^{-2} > c^{-2} + b^{-2}$ of Figure 3.6(d) by a rotation by 90° about the *c** axis. The three remaining

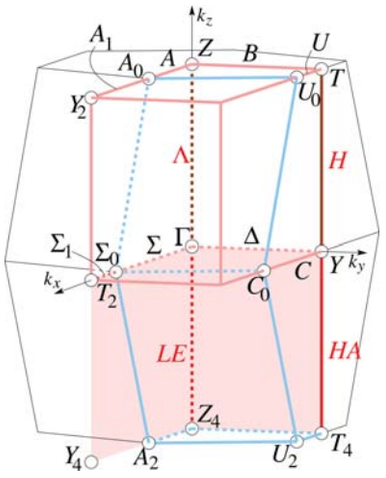


Fig. 1.5.5.7. Brillouin zone with asymmetric unit and representation domain of CDML for arithmetic crystal class *mm2F*: $a^{-2} > b^{-2} + c^{-2}$. Space groups $Fmm2 - C_{2v}^{18}$ (42), $Fdd2 - C_{2v}^{19}$ (43). Reciprocal-space group $(Imm2)^*$, No. 44: $a^{*2} > b^{*2} + c^{*2}$ (see Table 1.5.5.7). The representation domain of CDML is different from the asymmetric unit. Auxiliary points: T_4: $0, \frac{1}{2}, -\frac{1}{2}$; Y_4: $\frac{1}{2}, 0, -\frac{1}{2}$; Z_4: $0, 0, -\frac{1}{2}$. Flagpoles: $LE = [Z_4\,\Gamma]$ $0, 0, z$: $-\frac{1}{2} < z < 0$; $HA = [T_4\,Y]$ $0, \frac{1}{2}, z$: $-\frac{1}{2} < z < 0$. Wings: $JA \cup J_3 = [\Gamma\,T_2\,Y_4\,Z_4]$ $x, 0, z$: $0 < x < \frac{1}{2}, -\frac{1}{2} < z < 0$; $EA = [\Gamma\,Z_4\,T_4\,Y]$ $0, y, z$: $0 < y < \frac{1}{2}, -\frac{1}{2} < z < 0$.

Brillouin zones are displayed in Fig. 1.5.5.5 (see also Table 1.5.5.5), Fig. 1.5.5.6 (see also Table 1.5.5.6) and Fig. 1.5.5.7 (see also Table 1.5.5.7). Fig. 1.5.5.5 is a distorted cuboctahedron with 14 faces, 36 edges and 24 apices, 18 of which are visible. The Brillouin zones of Figs. 1.5.5.6 and 1.5.5.7 are distorted elongated rhombdodecahedra. There are 12 faces, 28 edges and 18 apices; 14 of them are visible.

1.5.5.4. Discussion

1.5.5.4.1. Representation domains and asymmetric units

When the symmetry of the reciprocal lattice allows, the shape of the asymmetric unit may be chosen to be much simpler than that of the representation domain.

Examples

(1) Arithmetic crystal class $4/mmmI$. The parameter ranges for the special lines and planes of the asymmetric unit and for general **k** vectors of the reciprocal-space group $(F4/mmm)^*$ [setting $(I4/mmm)^*$] are listed in Tables 1.5.5.3 and 1.5.5.4. One can describe the corresponding conditions of the representation domain by the boundary plane $x, y, z = \{1 + (c/a)^2[1 - 2(x + y)]\}/4$ which for $c/a < 1$ forms the triangle $[Z_0\,Z_2\,P]$ in Fig. 1.5.5.3 but for $c/a > 1$ the pentagon $[S_2\,R\,P\,G\,S]$ in Fig. 1.5.5.4. The inner points of this boundary plane are points of the general position *GP* with the exception of the line $Q = x, \frac{1}{2} - x, \frac{1}{4}$, which is a twofold rotation axis. The boundary conditions for the representation domain depend on c/a; they are much more complicated than those, $x, y, z = \frac{1}{4}$, for the asymmetric unit.

(2) Arithmetic crystal class *mm2F*, see Figs. 1.5.5.5 to 1.5.5.7. In the reciprocal-space group $(Imm2)^*$ the lines Λ and *LE* belong to Wintgen position 2 *a mm*2, as do the lines *Q*, *QA*, Λ_1 and Λ_3 if present. The lines *H* and *HA* belong to the Wintgen position 2 *b mm*2; as do the lines *G*, *GA*, H_1 and H_3 if present. The lines Σ, Σ_1, *A*, A_1, *C* and *U* belong to the plane $x, 0, z$; the lines Δ, *B*, B_1 and *D* belong to the plane $0, y, z$. The decisive boundary plane of the representation domain is $xa^{*2} + yb^{*2} + zc^{*2} = d^{*2}/4$, where $d^{*2} = a^{*2} + b^{*2} + c^{*2}$; it is a hexagon for Fig. 1.5.5.5 and a parallelogram for Figs. 1.5.5.6 and 1.5.5.7. There is no relation of the lattice parameters for which all the above-mentioned lines are realized on the surface of the representation domain simultaneously; either two or

Table 1.5.5.7. *List of* **k***-vector types for arithmetic crystal class mm2F:* $a^{-2} > b^{-2} + c^{-2}$

See Fig. 1.5.5.7. Parameter relations: $x = -\frac{1}{2}\alpha + \frac{1}{2}\beta + \frac{1}{2}\gamma$, $y = \frac{1}{2}\alpha - \frac{1}{2}\beta + \frac{1}{2}\gamma$, $z = \frac{1}{2}\alpha + \frac{1}{2}\beta - \frac{1}{2}\gamma$.

k-vector label, CDML	Wyckoff position of *IT* A, *cf.* Section 1.5.4.3				Parameters
Γ $0, 0, 0$	*ex*	2	*a*	*mm2*	$0, 0, 0$
Z $\frac{1}{2}, \frac{1}{2}, 0$	*ex*	2	*a*	*mm2*	$0, 0, \frac{1}{2}$
Λ $\alpha, \alpha, 0$	*ex*	2	*a*	*mm2*	$0, 0, z$: $0 < z < \frac{1}{2}$
LE $-\alpha, -\alpha, 0$	*ex*	2	*a*	*mm2*	$0, 0, z$: $-\frac{1}{2} < z < 0$
$\Gamma \cup Z \cup \Lambda \cup LE$		2	*a*	*mm2*	$0, 0, z$: $-\frac{1}{2} < z \leq \frac{1}{2}$
T $1, \frac{1}{2}, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, \frac{1}{2}$
Y $\frac{1}{2}, 0, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, 0$
H $\frac{1}{2} + \alpha, \alpha, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, z$: $0 < z < \frac{1}{2}$
HA $\frac{1}{2} - \alpha, -\alpha, \frac{1}{2}$	*ex*	2	*b*	*mm2*	$0, \frac{1}{2}, z$: $-\frac{1}{2} < z < 0$
$T \cup Y \cup H \cup HA$		2	*b*	*mm2*	$0, \frac{1}{2}, z$: $-\frac{1}{2} < z \leq \frac{1}{2}$
Σ $0, \alpha, \alpha$	*ex*	4	*c*	*.m.*	$x, 0, 0$: $0 < x \leq \sigma_0$
U $1, \frac{1}{2} + \alpha, \frac{1}{2} + \alpha$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, \frac{1}{2}$: $0 < x < u_0$
$U \sim \Sigma_1 = [\Sigma_0\,T_2]$					$x, 0, 0$: $\frac{1}{2} - u_0 = \sigma_0 < x < \frac{1}{2}$
A $\frac{1}{2}, \frac{1}{2} + \alpha, \alpha$	*ex*	4	*c*	*.m.*	$x, 0, \frac{1}{2}$: $0 < x < a_0$
C $\frac{1}{2}, \alpha, \frac{1}{2} + \alpha$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, 0$: $0 < x \leq c_0$
$C \sim A_1 = [A_0\,Y_2]$					$x, 0, \frac{1}{2}$: $a_0 = \frac{1}{2} - c_0 \leq x < \frac{1}{2}$
J $\alpha, \alpha + \beta, \beta$	*ex*	4	*c*	*.m.*	$x, 0, z$: $[\Gamma\,Z\,A_0\,\Sigma_0]$
JA $-\alpha, -\alpha + \beta, \beta$	*ex*	4	*c*	*.m.*	$x, 0, z$: $[\Gamma\,\Sigma_0\,A_2\,Z_4]$
K $\frac{1}{2} + \alpha, \alpha + \beta, \frac{1}{2} + \beta$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, z$: $[Y\,T\,U_0\,C_0]$
$K \sim J_3$					$x, 0, z$: $[T_2\,\Sigma_0\,A_2\,Y_4]$
KA $\frac{1}{2} - \alpha, -\alpha + \beta, \frac{1}{2} + \beta$	*ex*	4	*c*	*.m.*	$x, \frac{1}{2}, z$: $[Y\,C_0\,U_2\,T_4]$
$KA \sim J_1$					$x, 0, z$: $[T_2\,\Sigma_0\,A_0\,Y_2]$
$A \cup A_1 \cup J \cup J_1 \cup \Sigma \cup \Sigma_1 \cup JA \cup J_3$		4	*c*	*.m.*	$x, 0, z$: $0 < x < \frac{1}{2}, -\frac{1}{2} < z \leq \frac{1}{2}$
Δ $\alpha, 0, \alpha$	*ex*	4	*d*	*m..*	$0, y, 0$: $0 < y < \frac{1}{2}$
B $\frac{1}{2} + \alpha, \frac{1}{2}, \alpha$	*ex*	4	*d*	*m..*	$0, y, \frac{1}{2}$: $0 < y < \frac{1}{2}$
E $\alpha + \beta, \alpha, \beta$	*ex*	4	*d*	*m..*	$0, y, z$: $0 < y, z < \frac{1}{2}$
EA $-\alpha + \beta, -\alpha, \beta$	*ex*	4	*d*	*m..*	$0, y, z$: $0 < y < \frac{1}{2}, -\frac{1}{2} < z < 0$
$\Delta \cup B \cup E \cup EA$		4	*d*	*m..*	$0, y, z$: $0 < y < \frac{1}{2}, -\frac{1}{2} < z \leq \frac{1}{2}$
GP α, β, γ		8	*e*	1	x, y, z: $0 < x, y < \frac{1}{2}, 0 \leq z < \frac{1}{2}$

three of them do not appear and the length of the others depends on the boundary plane, see Tables and Figs. 1.5.5.5 to 1.5.5.7.

The boundary conditions for the asymmetric unit are independent of the lattice parameters and the boundary plane is always represented by the simple equation $x, y, \frac{1}{2}$: $0 < x, y < \frac{1}{2}$. By introducing flagpoles and wings, the description may become uni-arm.

1.5.5.4.2. Splitting of **k**-vector types

The Brillouin zone as well as the unit cell are always convex bodies; the same holds for the representation domain of CDML and for the choice of the asymmetric unit. It is thus sometimes unavoidable that the **k**-vector types are split and that the different parts belong to different arms and to different stars of **k** vectors. Sometimes this splitting of **k**-vector types may be avoided by an appropriate choice of the asymmetric unit; sometimes the introduction of flagpoles and wings is necessary to make the **k**-vector types uni-arm.

Examples

(1) In the reciprocal-space group $(\mathcal{G})^* = (Fm\bar{3}m)^*$, No. 225 of the arithmetic crystal class $m\bar{3}mI$ there are the lines of **k** vectors Λ (α, α, α) and F $(\frac{1}{2} - \alpha, -\frac{1}{2} + 3\alpha, \frac{1}{2} - \alpha)$ of CDML, p. 41. By Figure 1.5.5.1 one sees that the line Λ connects the points Γ and P, the line F connects the points P and H. One takes from the corresponding Table 1.5.5.1 the coefficients of $P = \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ and $H = 0, \frac{1}{2}, 0$. From these points or from the transformation listed at the top of Table 1.5.5.1 as 'Parameter relations' the coefficients of the line F are obtained as $F = x, \frac{1}{2} - x, x$; $0 < x < \frac{1}{4}$.

The inspection of the symmetry diagram of $Fm\bar{3}m$, No. 225, in *IT* A shows that a twofold rotation 2 (represented by the 4_2 $\frac{1}{4}, y, \frac{1}{4}$ screw-rotation axis) leaves the point P invariant, whereas the point H is mapped onto the point R $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$. More formally: the rotation is described by $x, \frac{1}{2} - x, x \rightarrow \frac{1}{2} - x, \frac{1}{2} - x, \frac{1}{2} - x$, where $0 < x < \frac{1}{4}$. The result is the line $F_1 = [R\,P]$. It is uni-arm to the line $\Lambda = x, x, x$ and the union $\Lambda \cup F_1$ forms the Wintgen position 32 *f* 3*m*. An analoguous result is obtained for the same lines in the arithmetic crystal class $m\bar{3}I$.

(2) In the following example the splitting of a Wintgen position happens if a representation domain of the Brillouin zone is chosen. The splitting can be avoided by the choice of the asymmetric unit. We consider the plane $x, y, 0$ in the arithmetic crystal class $4/mmmI$, see Fig. 1.5.5.4 and Table 1.5.5.4. In CDML this plane is split into the parts $C = [\Gamma\,S_2\,R\,X]$ and $D = [M\,S\,G] \sim [M_2\,S_2\,R]$. By the choice of the asymmetric unit the independent region of the Wintgen position is uni-arm: $[\Gamma M_2 X] = 16\ l\ m..\ x, y, 0$: $0 < x < y < \frac{1}{2}$.

(3) The splitting of a Wintgen position can be avoided if flagpoles and wings are admitted, *i.e.* if the minimal domain is described by a non-convex body. If one chooses in the first example of the arithmetic crystal classes $m\bar{3}mI$ and $m\bar{3}I$ the union $\Lambda \cup F_1$ for the line x, x, x, then $F_1 = [P\,R]$ forms a flagpole, whereas Λ forms an edge of the asymmetric unit, see Figs. 1.5.5.1 and 1.5.5.2.

The same holds for the Wintgen position 96 *k* ..*m* x, x, z of $m\bar{3}mI$. In the representation domain which is simultaneously the asymmetric unit, this Wintgen position is split into three parts *B*, *C* and *J*, which form three of the four walls of the (tetrahedral) minimal domain. By proper symmetry operations these three parts can be made uni-arm to the part *C*, such that their union $C \cup B_1 \cup J_1$ describes the independent part of that Wintgen position, see Fig. 1.5.5.1. The part *C* forms a wall of the asymmetric unit; the part $B_1 \cup J_1$ forms a wing, see Fig. 1.5.5.1.

1.5.5.4.3. **k**-vector types for non-holosymmetric space groups

The **k**-vector labels of CDML are primarily listed for the holosymmetric space groups. These lists are kept and supplemented for the non-holosymmetric space groups. In this way many superfluous **k**-vector labels are introduced.

Examples

(1) Arithmetic crystal class $m\bar{3}I$. In its reciprocal-space group $(Fm\bar{3})^*$, the introduction of the plane $AA = [\Gamma\,H_2\,N]$ is unnecessary because the plane $A = [\Gamma\,N\,H]$ of Wintgen position 96 *j* *m*.. of $(Fm\bar{3}m)^*$ can be extended to $A \cup AA = [\Gamma\,H_2\,H]$ in the reciprocal-space group $(Fm\bar{3})^*$, *cf.* Fig. 1.5.5.2 and Table 1.5.5.2. In $(Fm\bar{3})^*$, both planes, *A* and *AA*, belong to Wintgen position 48 *h* *m*... The parameter description is extended from $x, y, 0$: $0 < x < y < \frac{1}{2} - x (< \frac{1}{4})$ to $0 < y < \frac{1}{2} - x < \frac{1}{2}$.

(2) In the previous example, during the transition from the group $(Fm\bar{3}m)^*$ to the subgroup $(Fm\bar{3})^*$ the order of the little co-group of the special **k** vectors of $(Fm\bar{3}m)^*$ was not changed. In other cases, the little co-group may be reduced to a subgroup. Such **k** vectors may then be incorporated into a more general Wintgen position and described by an extension of the parameter range.

Arithmetic crystal class $m\bar{3}mI$, plane $[\Gamma\,H\,N] = x, y, 0$. In $(Fm\bar{3}m)^*$, see Fig. 1.5.5.1, all points (Γ, H, N) and lines (Δ, Σ, G) of the boundary of the asymmetric unit are special. In $(Fm\bar{3})^*$, see Fig. 1.5.5.2, the lines Δ and $[\Gamma\,H_2] \sim \Delta$ are special but Σ, *G* and $[N\,H_2] \sim G$ belong to the plane $(A \cup AA)$. The free parameter range on the line *G* is $0 < y < \frac{1}{4}$. Therefore, the parameter ranges of $(A \cup AA \cup G \cup \Sigma)$ in $x, y, 0$ can be taken as: $0 < x < \frac{1}{2} - y < \frac{1}{2}$ for $A \cup AA \cup \Sigma$ and $0 < y = \frac{1}{2} - x < \frac{1}{4}$ for *G*.

1.5.5.4.4. Ranges of independent parameters

In Section 1.5.4.3 a method for the determination of the parameter ranges was described. A few examples shall display the procedure.

(1) Arithmetic crystal class $m\bar{3}mI$, line $\Lambda \cup F_1$: In the reciprocal-space group $(Fm\bar{3}m)^*$ of the arithmetic crystal class $m\bar{3}mI$, the line x, x, x has stabilizer $\bar{3}m$ and little co-group $\overline{\mathcal{G}}^{\mathbf{k}} = 3m$. Therefore, the divisor is 12:6 = 2 and x is running from 0 to $\frac{1}{2}$.

The same result holds for the line $\Lambda \cup F_1$ in the reciprocal-space group $(Fm\bar{3})^*$ of the arithmetic crystal class $m\bar{3}I$: the stabilizer generated by $\bar{3}$ is of order 6, $|\overline{\mathcal{G}}^{\mathbf{k}}| = |\{\bar{3}\}| = 3$, the quotient is again $\frac{1}{2}$, the parameter range is the same as for $(Fm\bar{3}m)^*$.

(2) Arithmetic crystal class $m\bar{3}mI$, plane $B_1 \cup C \cup J_1$: In $(Fm\bar{3}m)^*$, the stabilizer of x, x, z is generated by *m*.*mm* and the centring translation $t(\frac{1}{2}, \frac{1}{2}, 0)$ mod (integer translations). They generate a group of order 16; $\overline{\mathcal{G}}^{\mathbf{k}}$ is ..*m* of order 2. The fraction of the plane is $\frac{2}{16} = \frac{1}{8}$ of the area $\sqrt{2}a^{*2}$ in the (centred) unit cell, as expressed by the parameter ranges $0 < x < \frac{1}{4}, 0 < z < \frac{1}{2}$. There are six arms of the star of x, x, z: x, x, z; $\bar{x}, x, z$; x, y, x; $x, y, \bar{x}$; x, y, y; $x, \bar{y}, y$. Three of them (x, x, z, $\bar{x}, x, z$ and x, y, x) are represented in the boundaries of the representation domain: $C = [\Gamma\,N\,P]$, $B = [H\,N\,P]$ and $J = [\Gamma\,H\,P]$, see Fig. 1.5.5.1. The areas of their parameter ranges are $\frac{1}{32}, \frac{1}{32}$ and $\frac{1}{16}$, respectively; the sum is $\frac{1}{8}$.

Arithmetic crystal class $m\bar{3}I$, the same result holds in the reciprocal-space group $(Fm\bar{3})^*$. The stabilizer generated by $2/m$.. and by the centring translation $t(\frac{1}{2}, \frac{1}{2}, 0)$ mod (integer translations) forms a group of order 8; the order of the little co-group $|\overline{\mathcal{G}}^{\mathbf{k}}| = |\{1\}| = 1$. The quotient is again $\frac{1}{8}$, the parameter range is the same as for $(Fm\bar{3}m)^*$ but the plane belongs to the general position *GP* because the little co-group is trivial.

(3) Arithmetic crystal class $m\bar{3}mI$, reciprocal-space group $(Fm\bar{3}m)^*$, plane $x, y, 0$: the stabilizer of the plane *A* is generated by $4/mmm$ and $t(\frac{1}{2}, \frac{1}{2}, 0)$, order 32, $\overline{\mathcal{G}}^{\mathbf{k}}$ (site-symmetry group) *m*..,

order 2. Consequently, $[\Gamma\ H\ N]$ is $\frac{1}{16}$ of the unit square a^{*2}: $0 < x < y < \frac{1}{2} - x$. In $(Fm\overline{3})^*$, the stabilizer of $x, y, 0$, here $A \cup AA$, is mmm. and $t(\frac{1}{2}, \frac{1}{2}, 0)$, order 16, with the same group $\overline{\mathcal{G}}^{\mathbf{k}} = m..$ of order 2. Therefore, $[\Gamma\ H_2\ H]$ is $\frac{1}{8}$ of the unit square a^{*2} in $(Fm\overline{3})^*$; $0 < y < \frac{1}{2} - x < \frac{1}{2}$.

(4) Arithmetic crystal class $m\overline{3}mI$, line $x, x, 0$: In $(Fm\overline{3}m)^*$ the stabilizer is generated by $m.mm$ and $t(\frac{1}{2}, \frac{1}{2}, 0)$ mod (integer translations), order 16, $\overline{\mathcal{G}}^{\mathbf{k}}$ is $m.2m$ of order 4. The divisor is 4 and thus $0 < x < \frac{1}{4}$. In $(Fm\overline{3})^*$ the stabilizer is generated by $2/m..$ and $t(\frac{1}{2}, \frac{1}{2}, 0)$ mod (integer translations), order 8, and $\overline{\mathcal{G}}^{\mathbf{k}} = m..$, order 2; the divisor is 4 again and $0 < x < \frac{1}{4}$ is restricted to the same range.

In the way just described the *inner* part of the parameter range can be fixed. The *boundaries* of the parameter range must be determined in addition:

(1) Arithmetic crystal classes $m\overline{3}mI$ and $m\overline{3}I$, *i.e.* $(Fm\overline{3}m)^*$ and $(Fm\overline{3})^*$, line x, x, x: The points $0, 0, 0$; $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$ (and $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$) are special points; the parameter ranges are open: $0 < x < \frac{1}{4}$, $\frac{1}{4} < x < \frac{1}{2}$.

(2) Arithmetic crystal class $m\overline{3}mI$, plane x, x, z: In $(Fm\overline{3}m)^*$ all corners Γ, N, N_2, H_2 and all edges are special points or lines. Therefore, the parameter ranges are open: x, x, z: $0 < x < \frac{1}{4}, 0 < z < \frac{1}{2}$, where the lines x, x, x: $0 < x < \frac{1}{4}$ and $x, x, \frac{1}{2} - x$: $0 < x < \frac{1}{4}$ are special lines and thus excepted.

(3) Arithmetic crystal classes $m\overline{3}mI$ and $m\overline{3}I$, plane $x, y, 0$: In both reciprocal-space groups, $(Fm\overline{3}m)^*$ and $(Fm\overline{3})^*$, $0 < x$ and $0 < y$ holds. The line $0, y, 0 = \Delta$ is a special line, its $\mathbf{k}$ vectors have little co-groups of higher order than that of the planes $x, y, 0$ and the boundaries of both planes are open. The same holds for the boundary $x, 0, 0 \sim 0, y, 0$ for $(Fm\overline{3})^*$. The $\mathbf{k}$ vectors of the lines $x, x, 0$ and $x, \frac{1}{2} - x, 0$, Σ and G, also have little co-groups of higher order and belong to other Wintgen positions in the representation domain (or asymmetric unit) of $(Fm\overline{3}m)^*$. Therefore, for the arithmetic crystal class $m\overline{3}mI$, the plane $A = x, y, 0$ is open at its boundaries $x, x, 0$ and $x, \frac{1}{2} - x, 0$ in the range $0 < x < \frac{1}{4}$. In the asymmetric unit of $(Fm\overline{3})^*$ the lines $x, x, 0$: $0 < x < \frac{1}{4}$ and $x, \frac{1}{2} - x, 0$: $0 < x < \frac{1}{4}$ belong to the plane, and the boundary of the plane A is here closed. The boundary line $x, \frac{1}{2} - x, 0$: $\frac{1}{4} < x < \frac{1}{2}$ of the plane AA is equivalent to the range $0 < x < \frac{1}{4}$ of the part A and thus does not belong to the asymmetric unit; here the boundary of the plane $A \cup AA$ is open.

1.5.6. Conclusions

International Tables for Crystallography Volume A can serve as a basis for the classification of irreps of space groups by using the concept of reciprocal-space groups. The main features of the crystallographic classification scheme are as follows.

(i) The asymmetric units of the conventional crystallographic unit cells are minimal domains of $\mathbf{k}$ space which are in many cases simpler than the representation domains of the Brillouin zones.

(ii) All $\mathbf{k}$-vector stars giving rise to the same type of irreps belong to the same Wintgen position and *vice versa*. They can be collected in one entry (uni-arm description) and are designated by the same Wintgen letter if flagpoles and wings are admitted.

(iii) The Wyckoff positions of *IT* A, interpreted as Wintgen positions, provide a complete list of the special $\mathbf{k}$ vectors in the Brillouin zone; the site symmetry of *IT* A is the little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$ of $\mathbf{k}$; the multiplicity per primitive unit cell is the number of arms of the star of $\mathbf{k}$. The Wintgen positions with 0, 1, 2, or 3 variable parameters correspond to special $\mathbf{k}$-vector points, $\mathbf{k}$-vector lines, $\mathbf{k}$-vector planes or to the set of all general $\mathbf{k}$ vectors, respectively.

The complete set of types of irreps is obtained by considering the irreps of one $\mathbf{k}$ vector per Wintgen position in the uni-arm description. A complete set of inequivalent irreps of $\mathcal{G}$ is obtained from these irreps by varying the parameters within their ranges.

Data on the independent parameter ranges are essential to make sure that exactly one $\mathbf{k}$ vector per orbit is represented in the representation domain Φ or in the asymmetric unit. Such data are often much easier to calculate for the asymmetric unit of the unit cell than for the representation domain of the Brillouin zone, in particular if a uni-arm description has been chosen, *cf.* Section 1.5.5. Such data can not be found in the cited tables of irreps.

The uni-arm description unmasks those $\mathbf{k}$ vectors which lie on the boundary of the Brillouin zone but belong to a Wintgen position which also contains inner $\mathbf{k}$ vectors, see the example of the lines Λ and F in $(Fm\overline{3}m)^*$ and $(Fm\overline{3})^*$. Such $\mathbf{k}$ vectors can not give rise to little-group representations obtained from projective representations of the little co-group $\overline{\mathcal{G}}^{\mathbf{k}}$.

The consideration of the basic domain Ω in relation to the representation domain Φ is unnecessary. It may even be misleading because special $\mathbf{k}$-vector subspaces of Ω frequently belong to more general types of $\mathbf{k}$ vectors in Φ. Space groups $\mathcal{G}$ with non-holohedral point groups can be referred to their reciprocal-space groups $(\mathcal{G})^*$ directly without reference to the types of irreps of the corresponding holosymmetric space group.

In principle both approaches are equivalent: the *traditional* one by Brillouin zone, basic domain and representation domain and the *crystallographic* one by unit cell and asymmetric unit. Moreover, it is not difficult to relate one approach to the other, see Figs. and Tables 1.5.5.1 to 1.5.5.7. The conclusions show that the crystallographic approach for the description of irreps of space groups has several advantages as compared with the traditional approach. Owing to these advantages, CDML have already accepted the crystallographic approach for triclinic and monoclinic space groups. However, the advantages are not restricted to such low symmetries. In particular, the simple boundary conditions and shapes of the asymmetric units result in simple equations for the boundaries and shapes of volume elements and facilitate numerical calculations, integrations *etc.* If there are special reasons to prefer $\mathbf{k}$ vectors inside or on the boundary of the Brillouin zone to those outside, then the advantages and disadvantages of both approaches have to be compared in order to find the optimal way to solve the problem.

The crystallographic approach may be realized in three different ways:

(1) In the *uni-arm description* one lists each $\mathbf{k}$-vector star exactly once by indicating the parameter field of the representing $\mathbf{k}$ vector. Advantages are the transparency of the presentation and the relatively small effort for the derivation of the list. A disadvantage may be that there are protruding flagpoles or wings. Points of flagpoles or wings are no longer neighbours of inner points (an inner point has a full three-dimensional sphere of neighbours which belong to the asymmetric unit).

(2) In the *compact description* one lists each $\mathbf{k}$ vector exactly once such that each point of the asymmetric unit is either an inner point itself or has inner points as neighbours. Such a description may not be uni-arm for some Wintgen positions, and the determination of the parameter ranges may become less straightforward.

(3) In the *non-unique description* one gives up the condition that each $\mathbf{k}$ vector is listed exactly once. The uni-arm and the compact descriptions are combined but the equivalence relations ($\sim$) are stated explicitly for those $\mathbf{k}$ vectors which occur in more than one entry. Such tables are most informative and not too complicated for practical applications.

The authors wish to thank the editor of this volume, Uri Shmueli, for his patient support, for his encouragement and for his valuable help. They are grateful to the former Chairman of the Commission on *International Tables*, Theo Hahn, for his interest and advice. Part of the material in this chapter was first published as an article of the same title in *Z. Kristallogr.* (1995),

210, 243–254. We are indebted to R. Oldenbourg Verlag, Munich, Germany, for allowing us to reprint parts of this article.

APPENDIX A1.5.1
Reciprocal-space groups $(\mathcal{G})^*$

This table is based on Table 1 of Wintgen (1941).

In order to obtain the Hermann–Mauguin symbol of $(\mathcal{G})^*$ from that of $\mathcal{G}$, one replaces any screw rotations by rotations and any glide reflections by reflections. The result is the symmorphic space group $\mathcal{G}_0(\mathcal{G})$. For most space groups $\mathcal{G}$, the reciprocal-space group $(\mathcal{G})^*$ is isomorphic to $\mathcal{G}_0(\mathcal{G})$, *i.e.* $\mathcal{G}_0(\mathcal{G})$ and $(\mathcal{G})^*$ belong to the same arithmetic crystal class. In the following cases $(\mathcal{G})^*$ is isomorphic to a symmorphic space group $\mathcal{G}_0$ which is different from $\mathcal{G}_0(\mathcal{G})$. Thus the arithmetic crystal classes of $\mathcal{G}$ and $(\mathcal{G})^*$ are different, *i.e.* $(\mathcal{G})^*$ can not be obtained in this simple way:

(1) If the lattice symbol of $\mathcal{G}$ is F or I, it has to be replaced by I or F, *e.g.* $(\mathcal{G})^*$ is isomorphic to $Imm2$ for the arithmetic crystal class $\mathcal{G} = mm2F$. The tetragonal space groups form an exception to this rule; for these the symbol I persists.

(2) The other exceptions are listed in the following table (for the symbols of the arithmetic crystal classes see *IT* A, Section 8.2.3):

Arithmetic crystal class of $\mathcal{G}$	Reciprocal-space group $(\mathcal{G})^*$
$\bar{4}m2I$	$I\bar{4}2m$
$\bar{4}2mI$	$I\bar{4}m2$
$321P$	$P312$
$312P$	$P321$
$3m1P$	$P31m$
$31mP$	$P3m1$
$\bar{3}1mP$	$P\bar{3}m1$
$\bar{3}m1P$	$P\bar{3}1m$
$\bar{6}m2P$	$P\bar{6}2m$
$\bar{6}2mP$	$P\bar{6}m2$

References

Altmann, S. L. (1977). *Induced Representations in Crystals and Molecules.* London: Academic Press.

Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. (2006). *Bilbao Crystallographic Server. II. Representations of crystallographic point groups and space groups. Acta Cryst.* A**62**, 115–128.

Aroyo, M. I., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006). *Bilbao Crystallographic Server: I. Databases and crystallographic computing programs. Z. Kristallogr.* **221**, 15–27.

Aroyo, M. I. & Wondratschek, H. (1995). *Crystallographic viewpoints in the classification of space-group representations. Z. Kristallogr.* **210**, 243–254.

Bilbao Crystallographic Server (1998). http://www.cryst.ehu.es/.

Bouckaert, L. P., Smoluchowski, R. & Wigner, E. P. (1936). *Theory of Brillouin zones and symmetry properties of wave functions in crystals. Phys. Rev.* **50**, 58–67.

Boyle, L. L. (1986). *The classification of space group representations.* In *Proceedings of the 14th International Colloquium on Group-Theoretical Methods in Physics*, pp. 405–408. Singapore: World Scientific.

Bradley, C. J. & Cracknell, A. P. (1972). *The Mathematical Theory of Symmetry in Solids: Representation Theory for Point Groups and Space Groups.* Oxford: Clarendon Press.

Cracknell, A. P., Davies, B. L., Miller, S. C. & Love, W. F. (1979). *Kronecker Product Tables*, Vol. 1, *General Introduction and Tables of Irreducible Representations of Space Groups.* New York: IFI/Plenum.

Davies, B. L. & Cracknell, A. P. (1976). *Some comments on and addenda to the tables of irreducible representations of the classical space groups published by S. C. Miller and W. F. Love. Acta Cryst.* A**32**, 901–903.

Davies, B. L. & Dirl, R. (1987). *Various classification schemes for irreducible space group representations.* In *Proceedings of the 15th International Colloquium on Group-Theoretical Methods in Physics*, edited by R. Gilmore, pp. 728–733. Singapore: World Scientific.

Delaunay, B. (1933*a*). *Neue Darstellung der geometrischen Kristallographie. Z. Kristallogr.* **84**, 109–149.

Delaunay, B. (1933*b*). *Berichtigung zur Arbeit "Neue Darstellung der geometrischen Kristallographie". Z. Kristallogr.* **85**, 332.

International Tables for Crystallography (2005). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn, 5th ed. Heidelberg: Springer.

Jan, J.-P. (1972). *Space groups for Fermi surfaces. Can. J. Phys.* **50**, 925–927.

Jansen, L. & Boon, M. (1967). *Theory of Finite Groups. Applications in Physics: Symmetry Groups of Quantum Mechanical Systems.* Amsterdam: North-Holland.

Janssen, T. (2003). *International Tables for Crystallography*, Vol. D, *Physical Properties of Crystals*, edited by A. Authier, ch. 1.2, *Representations of crystallographic groups.* Dordrecht: Kluwer Academic Publishers.

Kovalev, O. V. (1986). *Irreducible and Induced Representations and Co-representations of Fedorov Groups.* Moscow: Nauka.

Lomont, J. S. (1959). *Applications of Finite Groups.* New York: Academic Press.

Miller, S. C. & Love, W. F. (1967). *Tables of Irreducible Representations of Space Groups and Co-representations of Magnetic Space Groups.* Boulder: Pruett Press.

Raghavacharyulu, I. V. V. (1961). *Representations of space groups. Can. J. Phys.* **39**, 830–840.

Rosen, J. (1981). *Resource letter SP-2: Symmetry and group theory in physics. Am. J. Phys.* **49**, 304–319.

Slater, L. S. (1962). *Quantum Theory of Molecules and Solids*, Vol. 2. Amsterdam: McGraw-Hill.

Smith, J. V. (1982). *Geometrical and Structural Crystallography*. New York: John Wiley & Sons.

Stokes, H. T. & Hatch, D. M. (1988). *Isotropy Subgroups of the 230 Crystallographic Space Groups.* Singapore: World Scientific.

Stokes, H. T., Hatch, D. M. & Nelson, H. M. (1993). *Landau, Lifshitz, and weak Lifshitz conditions in the Landau theory of phase transitions in solids. Phys. Rev. B*, **47**, 9080–9083.

Wintgen, G. (1941). *Zur Darstellungstheorie der Raumgruppen. Math. Ann.* **118**, 195–215.

Zak, J., Casher, A., Glück, M. & Gur, Y. (1969). *The Irreducible Representations of Space Groups.* New York: Benjamin.

International Tables for Crystallography (2010). Vol. B, Chapter 2.1, pp. 195–214.

2.1. Statistical properties of the weighted reciprocal lattice

By U. Shmueli and A. J. C. Wilson†

2.1.1. Introduction

The structure factor of the *hkl* reflection is given by

$$F(hkl) = \sum_{j=1}^{N} f_j \exp[2\pi i(hx_j + ky_j + lz_j)], \qquad (2.1.1.1)$$

where f_j is the atomic scattering factor [complex if there is appreciable dispersion; see Chapter 1.2 and *IT* C (2004, Section 4.2.6)], $x_j y_j z_j$ are the fractional coordinates of the jth atom and N is the number of atoms in the unit cell. The present chapter is concerned with the statistical properties of the structure factor F and the intensity $I = FF^*$, such as their average values, variances, higher moments and their probability density distributions.

Equation (2.1.1.1) expresses F as a function of two conceptually different sets of variables: *hkl* taking on integral values in reciprocal space and *xyz* in general having non-integral values in direct space, although the special positions tabulated for each space group in *IT* A (2005) may include the integers 0 and 1. In special positions, the non-integers often include rational fractions, but in general positions they are in principle irrational. Although *hkl* and *xyz* appear to be symmetrical variables in (2.1.1.1), these limitations on their values mean that one can consider two different sets of statistical properties. In the first we seek, for example, the average intensity of the *hkl* reflection (indices fixed) as the positional parameters of the N atoms are distributed with equal probability over the continuous range 0–1. In the second, we seek, for example, the average intensity of the observable reflections (or of a subgroup of them having about the same value of $\sin\theta$) with the values of *xyz* held constant at the values they have, or are postulated to have, in a crystal structure. Other examples are obtained by substituting the words 'probability density' for 'average intensity'. For brevity, we may call the statistics resulting from the first process *fixed-index* (continuously variable parameters being understood), and those resulting from the second process *fixed-parameter* (integral indices being understood). Theory based on the first process is (comparatively) easy; theory based on the second hardly exists, although there is a good deal of theory concerning the conditions under which the two processes will lead to the same result (Hauptman & Karle, 1953; Giacovazzo, 1977, 1980). Mathematically, of course, the condition is that the phase angle

$$\vartheta = 2\pi(hx + ky + lz) \qquad (2.1.1.2)$$

should be distributed with uniform probability over the range 0–2π, whichever set of variables is regarded as fixed, but it is not clear when this distribution can be expected in practice for fixed-parameter averaging. The usual conclusion is that the uniform distribution will be realized if there are enough atoms, if the atomic coordinates do not approximate to rational fractions, if there are enough reflections and if stereochemical effects are negligible (Shmueli *et al.*, 1984).

Obviously, the second process (fixed parameters, varying integral indices) corresponds to the observable reality, and various approximations to it have been attempted, in preference to assuming its equivalence with the first. For example, a third (approximate) method of averaging has been used (Wilson, 1949, 1981): *xyz* are held fixed and *hkl* are treated as continuous variables.

† Deceased.

2.1.2. The average intensity of general reflections

2.1.2.1. Mathematical background

The process may be illustrated by evaluating, or attempting to evaluate, the average intensity of reflection by the three processes. The intensity of reflection is given by multiplying equation (2.1.1.1) by its complex conjugate:

$$I = FF^* \qquad (2.1.2.1)$$

$$= \sum_{j,k} f_j f_k^* \exp\{2\pi i[h(x_j - x_k) + \ldots]\} \qquad (2.1.2.2)$$

$$= \Sigma + \sum_{j\neq k} f_j f_k^* \exp\{2\pi i[h(x_j - x_k) + \ldots]\}, \qquad (2.1.2.3)$$

where

$$\Sigma = \sum_j f_j f_j^* \qquad (2.1.2.4)$$

is the sum of the squares of the moduli of the atomic scattering factors. Wilson (1942) argued, without detailed calculation, that the average value of the exponential term would be zero and hence that

$$\langle I\rangle = \Sigma. \qquad (2.1.2.5)$$

Averaging equation (2.1.2.3) for *hkl* fixed, *xyz* ranging uniformly over the unit cell – the first process described above – gives this result identically, without complication or approximation. Ordinarily the second process cannot be carried out. We can, however, postulate a special case in which it is possible. We take a homoatomic structure and before averaging we correct the f's for temperature effects and the fall-off with $\sin\theta$, so that ff^* is the same for all the atoms and is independent of *hkl*. If the range of *hkl* over which the expression for I has to be averaged is taken as a parallelepiped in reciprocal space with h ranging from $-H$ to $+H$, k from $-K$ to $+K$, l from $-L$ to $+L$, equation (2.1.2.2) can be factorized into the product of the sums of three geometrical progressions. Algebraic manipulation then easily leads to

$$\langle I\rangle = ff^* \sum_j \sum_k \frac{\sin\pi N_H(x_j - x_k)}{N_H \sin\pi(x_j - x_k)} \times \frac{\sin\pi N_K(y_j - y_k)}{N_K \sin\pi(y_j - y_k)} \frac{\sin\pi N_L(z_j - z_k)}{N_L \sin\pi(z_j - z_k)}, \qquad (2.1.2.6)$$

where $N_H = 2H + 1$, $N_K = 2K + 1$ and $N_L = 2L + 1$. The terms with $j = k$ give Σ, but the remaining terms are not zero. Because of the periodic nature of the trigonometric terms, the effective coordinate differences are never greater than 0.5 and in a structure of any complexity there will be many much less than 0.5. For $HKL = 000$, in fact, $\langle I\rangle$ becomes the square of the modulus of the sum of the atomic scattering factors,

$$\langle I\rangle = \Phi\Phi^*, \qquad (2.1.2.7)$$

where

$$\Phi = \sum_{j=1}^{N} f_j, \qquad (2.1.2.8)$$

and not the sum of the squares of their moduli; for larger HKL, $\langle I\rangle$ rapidly decreases to Σ and then oscillates about that value.

Copyright © 2010 International Union of Crystallography

Wilson (1949, especially Section 2.1.1) suggested that the regions of averaging should be chosen so that at least one index of every reflection is ≥ 2 if $\langle I \rangle$ is to be identified with Σ, and this has proven to be a useful rule-of-thumb.

The third process of averaging replaces the sum over integral values of the indices by an integration over continuous values, the appropriate values of the limits in this example being $-(H + 1/2)$ to $+(H + 1/2)$. The effect is to replace the sines in the denominators, but not in the numerators, of equation (2.1.2.6) by their arguments, and this is equivalent to the approximation $\sin x \simeq x$ in the denominators only. This is a good approximation for atoms close together in the structure and thus giving the largest terms in the sums in equation (2.1.2.6), and gives the correct sign and order of magnitude even for x having its maximum value of $\pi/2$.

2.1.2.2. Physical background

The preceding section has used mathematical arguments. From a physical point of view, the radiation diffracted by atoms that are resolved will interfere destructively, so that the resulting intensity will be the sum of the intensities diffracted by individual atoms, whereas that from completely unresolved atoms will interfere constructively, so that *amplitudes* rather than intensities add. In intermediate cases there will be partial constructive interference. Resolution in accordance with the Rayleigh (1879) criterion requires that $s = (2 \sin \theta)/\lambda$ should be greater than half the reciprocal of the minimum interatomic distance in the crystal (Wilson, 1979); full resolution requires a substantial multiple of this. This criterion is essentially equivalent to that proposed from the study of a special case of the second process in the preceding section.

2.1.2.3. An approximation for organic compounds

In organic compounds there are very many interatomic distances of about 1.5 or 1.4 Å. Adoption of the preceding criterion would mean that the inner portion of the region of reciprocal space accessible by the use of copper $K\alpha$ radiation is not within the sphere of intensity statistics based on fixed-index (first process) averaging. No substantial results are available for fixed-parameter (second process) averaging, and very few from the approximation to it (third process).

To the extent to which the third process is acceptable, an approximation to the variation of $\langle I \rangle$ with $\sin \theta$ is obtainable. The exponent in equation (2.1.2.2) can be written as

$$2\pi i s r_{jk} \cos \psi, \tag{2.1.2.9}$$

where s is the radial distance in reciprocal space, r_{jk} is the distance from the jth to the kth atom and ψ is the angle between the vectors **s** and **r**. Averaging over a sphere of radius s, with ψ treated as the colatitude, gives

$$\langle I \rangle = \sum_j \sum_k f_j f_k \frac{\sin 2\pi s r_{jk}}{2\pi s r_{jk}}. \tag{2.1.2.10}$$

This is the familiar Debye expression. It has the correct limits for s zero and s large, and is in accord with the argument from resolution.

2.1.2.4. Effect of centring

In the preceding discussion there has been a tacit assumption that the lattice is primitive. A centred crystal can always be referred to a primitive lattice and if this is done no change is required. If the centred lattice is retained, many reflections are identically zero and the intensity of the nonzero reflections is enhanced by a factor of two (I and C lattices) or four (F lattice), so that the average intensity of all the reflections, zero and nonzero taken together, is unchanged.

Table 2.1.3.1. *Intensity-distribution effects of symmetry elements and centred lattices causing systematic absences*

Abbreviations and orientation of axes: A = acentric distribution, C = centric distribution, Z = systematically zero, S = distribution parameter, $\langle I \rangle$ = average intensity. Axes are parallel to **c**, planes are perpendicular to **c**.

Element	Reflections	Distribution	S/Σ	$\langle I \rangle/\Sigma$
2_1	*hkl*	A	1	1
	*hk*0	C	1	1
	00*l*	$(Z + A)/2$	1	2
$3_1, 3_2$	*hkl*	A	1	1
	*hk*0	A	1	1
	00*l*	$(2Z + A)/3$	1	3
$4_1, 4_3$	*hkl*	A	1	1
	*hk*0	C	1	1
	00*l*	$(3Z + A)/4$	1	4
4_2	*hkl*	A	1	1
	*hk*0	C	1	1
	00*l*	$(Z + A)/2$	2	4
$6_1, 6_5$	*hkl*	A	1	1
	*hk*0	C	1	1
	00*l*	$(5Z + A)/6$	1	6
$6_2, 6_4$	*hkl*	A	1	1
	*hk*0	C	1	1
	00*l*	$(2Z + A)/3$	2	6
6_3	*hkl*	A	1	1
	*hk*0	C	1	1
	00*l*	$(Z + A)/2$	3	6
a	*hkl*	A	1	1
	*hk*0	$(Z + A)/2$	1	2
	00*l*	C	1	1
	0*k*0	A	2	2
C, I	All	$(Z + A)/2$	1	2
F	All	$(3Z + A)/2$	1	4

Other symmetry elements affect only zones and rows of reflections, and so do not affect the general average when the total number of reflections is large. Their effect on zones and rows is discussed in Section 2.1.3.

2.1.3. The average intensity of zones and rows

2.1.3.1. Symmetry elements producing systematic absences

Symmetry elements can be divided into two types: those that cause systematic absences and those that do not. Those producing systematic absences (glide planes and screw axes) produce at the same time groups of reflections (confined to zones and rows in reciprocal space, respectively) with an average intensity an integral[1] multiple of the general average. The effects for single symmetry elements of this type are given in Table 2.1.3.1 for the general reflections *hkl* and separately for any zones and rows that are affected. The 'average multipliers' are given in the column headed $\langle I \rangle/\Sigma$; 'distribution' and 'distribution parameters' are treated in Section 2.1.5. As for the centring, the fraction of reflections missing and the integer multiplying the average are related in such a way that the overall intensity is unchanged. The mechanism for compensation for the reflections with enhanced intensity is obvious.

2.1.3.2. Symmetry elements not producing systematic absences

Certain symmetry elements not producing absences (mirror planes and rotation axes) cause equivalent atoms to coincide in a

[1] The multiple is given as an exact integer for fixed-index averaging, an approximate integer for fixed-parameter averaging. Statements should be understood to refer to fixed-index averaging unless the contrary is explicitly stated.

Table 2.1.3.2. *Intensity-distribution effects of symmetry elements not causing systematic absences*

Abbreviations and orientation of axes: A = acentric distribution, C = centric distribution, S = distribution parameter, $\langle I\rangle$ = average intensity. Axes are parallel to **c**, planes are perpendicular to **c**.

Element	Reflections	Distribution	$S/\Sigma = \langle I\rangle/\Sigma$
1	All	A	1
$\bar{1}$	All	C	1
2	hkl	A	1
	$hk0$	C	1
	$00l$	A	2
$\bar{2} = m$	hkl	A	1
	$hk0$	A	2
	$00l$	C	1
3	hkl	A	1
	$hk0$	A	1
	$00l$	A	3
$\bar{3}$	hkl	C	1
	$hk0$	C	1
	$00l$	C	3
4	hkl	A	1
	$hk0$	C	1
	$00l$	A	4
$\bar{4}$	hkl	A	1
	$hk0$	C	1
	$00l$	C	2
6	hkl	A	1
	$hk0$	C	1
	$00l$	A	6
$\bar{6} = 3/m$	hkl	A	1
	$hk0$	A	2
	$00l$	C	3

plane or a line projection and hence produce a zone or row in reciprocal space for which the average intensity is an integral multiple of the general average (Wilson, 1950); the effects of single such symmetry elements are given in Table 2.1.3.2. There is, however, no obvious mechanism for compensation for this enhancement. When reflections are few this may be an important matter in assigning an approximate absolute scale by comparing observed and calculated intensities. Wilson (1964), Nigam (1972) and Nigam & Wilson (1980), noting that in such cases the finite size of atoms results in forbidden ranges of positional parameters, have shown that there is a diminution of the intensity of layers (rows) in the immediate neighbourhood of the enhanced zones (rows), just sufficient to compensate for the enhancement. In forming general averages, therefore, reflections from enhanced zones or rows should be included at their full intensity, not divided by the multiplier; the matter is discussed in more detail by Wilson (1987*a*). It should be noted, however, that organic structures containing molecules related by rotation axes are rare, and such structures related by mirror planes are even rarer (Wilson, 1993).

2.1.3.3. More than one symmetry element

Further alterations of the intensities occur if two or more such symmetry elements are present in the space group. The effects were treated in detail by Rogers (1950), who used them to construct a table for the determination of space groups by supplementing the usual knowledge of Laue group with statistical information. Only two pairs of space groups, the orthorhombic $I222$ and $I2_12_12_1$, and their cubic supergroups $I23$ and $I2_13_1$, remained unresolved. Examination of this table shows that what statistical information does is to resolve the Laue group into point groups; the further resolution into space groups is equivalent to the use of Table 3.1.4.1 in *IT* A (2005). The statistical consequences of each point group, as given by Rogers, are reproduced in Table 2.1.3.3.

Table 2.1.3.3. *Average multiples for the 32 point groups (modified from Rogers, 1950)*

The multiple gives S/Σ for the row and zone corresponding to the principal axis of the point-group symbol; those for the secondary and tertiary axes are given when the symbol contains such axes.

	Principal		Secondary		Tertiary	
Point group	Row	Zone	Row	Zone	Row	Zone
1	1	1				
$\bar{1}$	1	1				
2	2	1				
m	1	2				
$2/m$	2	2				
222	2	1	2	1	2	1
$mm2$	2	2	2	2	4	1
mmm	4	2	4	2	4	2
4	4	1				
$\bar{4}$	2	1				
$4/m$	4	2				
422	4	1†	2	1	2	1
$4mm$	8	1	2	2	2	2
$\bar{4}2m$	4	1	2	1	2	2
$4/mmm$	8	2	4	2	4	2
3	3	1				
$\bar{3}$	3	1				
321	3	1	2	1	1	1
$3m1$	6	1	1	2	2	
$31m$	6	1	2	2	2	1
6	6	1				
$\bar{6}$	3	2				
$6/m$	6	2				
622	6	1	2	1	2	1
$6mm$	12	1	2	2	2	2
$\bar{6}m2$	6	2	2	2	4	1
$6/mmm$	12	2	4	2	4	2
231	2	1	3	1	1	1
$m\bar{3}1$	4	2	3	1	1	1
432	4	1	3	1	2	1
$\bar{4}3m$	4	1	6	1	2	2
$m\bar{3}m$	8	2	6	2	4	2

Note. The pairs of point groups, 1 and $\bar{1}$ and 3 and $\bar{3}$, not distinguished by average multiples, may be distinguished by their centric and acentric probability density functions. † The entry for the principal zone for the point group 422 was given incorrectly as 2 in the first edition of this volume.

2.1.4. Probability density distributions – mathematical preliminaries

For the purpose of this chapter, 'ideal' probability distributions or probability density functions are the asymptotic forms obtained by the use of the central-limit theorem when the number of atoms in the unit cell, N, is sufficiently large. In order to derive them it is necessary to outline the properties of characteristic functions and to state alternative conditions for the validity of the central-limit theorem; the distributions themselves are derived in Section 2.1.5.

2.1.4.1. Characteristic functions

The average value of $\exp(itx)$ is very important in probability theory; it is called the characteristic function of the distribution $f(x)$ and is denoted by $C_x(t)$ or, when no confusion can arise, by

$C(t)$. It exists for all legitimate distributions, whether discrete or continuous. In the continuous case it is given by

$$C(t) = \int_{-\infty}^{\infty} \exp(itx) f(x)\,\mathrm{d}x, \tag{2.1.4.1}$$

and is thus the Fourier transform of $f(x)$. In many cases it can be obtained from known integrals. For example, for the Cauchy distribution,

$$C(t) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{\exp(itx)}{a^2 + x^2}\,\mathrm{d}x \tag{2.1.4.2}$$

$$= \exp(-a|t|), \tag{2.1.4.3}$$

and for the normal distribution,

$$C(t) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right) \exp(itx)\,\mathrm{d}x \tag{2.1.4.4}$$

$$= \exp\left(imt - \frac{\sigma^2 t^2}{2}\right). \tag{2.1.4.5}$$

Since the characteristic function is the Fourier transform of the distribution function, the converse is true, and if the characteristic function is known the probability distribution function can be obtained by the use of Fourier inversion theorem,

$$f(x) = (1/2\pi) \int_{-\infty}^{\infty} \exp(-itx) C(t)\,\mathrm{d}t. \tag{2.1.4.6}$$

An alternative approach to the derivation of the distribution from a known characteristic function will be discussed below.

The most important property of characteristic functions in crystallography is the following: if x and y are independent random variables with characteristic functions $C_x(t)$ and $C_y(t)$, the characteristic function of their sum

$$z = x + y \tag{2.1.4.7}$$

is the product

$$C_z(t) = C_x(t) C_y(t). \tag{2.1.4.8}$$

Obviously this can be extended to any number of independent random variables.

When the moments exist, the characteristic function can be expanded in a power series in which the kth term is $m_k(it)^k/k!$. If the power series

$$\exp(itx) = 1 + itx + \frac{(it)^2 x^2}{2!} + \frac{(it)^3 x^3}{3!} + \ldots \tag{2.1.4.9}$$

is substituted in equation (2.1.4.1), one obtains

$$C(t) = 1 + itm_1' + \frac{(it)^2 m_2'}{2!} + \frac{(it)^3 m_3'}{3!} + \ldots. \tag{2.1.4.10}$$

The moments are written with primes in order to indicate that equation (2.1.4.10) is valid for moments about an arbitrary origin as well as for moments about the mean. If the random variable is transformed by a change of origin and scale, say

$$y = \frac{x - a}{b}, \tag{2.1.4.11}$$

the characteristic function for y becomes

$$C_y(t) = b \exp(-iat/b) C_x(t). \tag{2.1.4.12}$$

2.1.4.2. The cumulant-generating function

A function that is often more useful than the characteristic function is its logarithm, the cumulant-generating function:

$$K(t) = \log C(t) = k_1 + \frac{k_2(it)^2}{2!} + \frac{k_3(it)^3}{3!} + \ldots, \tag{2.1.4.13}$$

where the k's are called the *cumulants* and may be regarded as being defined by the equation. They can be evaluated in terms of the moments by combining the series (2.1.4.10) for $C(t)$ with the ordinary series for the logarithm and equating the coefficients of t^r. In most cases the process as described is tedious, but it can be shortened by use of a general method [Stuart & Ord (1994), Section 3.14, pp. 87–88; Exercise 3.19, p. 119]. Obviously, the cumulants exist only if the moments exist. The first few relations are

$$\begin{aligned} k_0 &= 0 \\ k_1 &= m_1' \\ k_2 &= m_2 = m_2' - (m_1')^2 \\ k_3 &= m_3 = m_3' - 3m_2'm_1' + 2(m_1')^2 \\ k_4 &= m_4 - 3(m_2)^2 \\ &= m_4' - m_3'm_1' - 3(m_2')^2 + 12m_2'(m_1)^2 - 6(m_1')^4. \end{aligned} \tag{2.1.4.14}$$

Such expressions and their converses up to k_{10} are given by Stuart & Ord (1994, pp. 88–91). Since all the cumulants except k_1 can be expressed in terms of the central moments only (*i.e.*, those unprimed), only k_1 is changed by a change of the origin. Because of this property, they are sometimes called the semi-invariants (or seminvariants) of the distribution. Since addition of random variables is equivalent to the multiplication of their characteristic functions [equation (2.1.4.8)] and multiplication of functions is equivalent to the addition of their logarithms, each cumulant of the distribution of the sum of a number of random variables is equal to the sum of the cumulants of the distribution functions of the individual variables – hence the name cumulants. Although the cumulants (except k_1) are independent of a change of origin, they are not independent of a change of scale. As for the moments, a change of scale simply multiplies them by a power of the scale factor; if $y = x/b$

$$(k_y)_r = (k_x)_r / b^r. \tag{2.1.4.15}$$

The cumulants of the normal distribution are particularly simple. From equation (2.1.4.5), the cumulant-generating function of a normal distribution is

$$K(t) = imt - \sigma^2 t^2/2 \tag{2.1.4.16}$$

$$k_1 = m \tag{2.1.4.17}$$

$$k_2 = \sigma^2, \tag{2.1.4.18}$$

all cumulants with $r > 2$ are identically zero.

2.1.4.3. The central-limit theorem

A simple form of this important theorem can be stated as follows:

If $x_1, x_2, \ldots, x_n$ are independent and identically distributed random variables, each of them having the same mean m and variance σ^2, then the sum

$$S_n = \sum_{j=1}^{n} x_j \tag{2.1.4.19}$$

tends to be normally distributed – independently of the distribution(s) of the individual random variables – with mean nm and variance $n\sigma^2$, provided n is sufficiently large.

In order to prove this theorem, let us define a standardized random variable corresponding to the sum S_n, *i.e.*, such that its mean is zero and its variance is unity:

$$\hat{S}_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\sum_{j=1}^{n}(x_j - m)}{\sigma\sqrt{n}} \equiv \sum_{j=1}^{n}\frac{W_j}{\sqrt{n}}, \qquad (2.1.4.20)$$

where $W_j = (x_j - m)/\sigma$ is a standardized single random variable. The characteristic function of $\hat{S}_n$ is therefore given by

$$\begin{aligned} C_n(\hat{S}_n, t) &= \langle \exp(it\hat{S}_n)\rangle \\ &= \left\langle \exp\left[it\sum_{j=1}^{n}\frac{W_j}{\sqrt{n}}\right]\right\rangle \qquad (2.1.4.21)\\ &= \prod_{j=1}^{n}\left\langle \exp\left[it\frac{W_j}{\sqrt{n}}\right]\right\rangle \qquad (2.1.4.22)\\ &= \left\{\left\langle \exp\left[it\frac{W_1}{\sqrt{n}}\right]\right\rangle\right\}^n, \qquad (2.1.4.23)\end{aligned}$$

where the brackets $\langle\ \rangle$ denote the operation of averaging with respect to the appropriate probability density function (p.d.f.) [*cf.* equation (2.1.4.1)]. Equation (2.1.4.22) follows from equation (2.1.4.21) by the assumption of independence, while the assumption of identically distributed variables leads to the identity of the characteristic functions of the individual variables – as seen in equation (2.1.4.23).

On the assumption that moments of all the orders exist – a most plausible assumption in situations usually encountered in structure-factor statistics – we can now expand the characteristic function of a single variable in a power series [*cf.* equation (2.1.4.10)]:

$$\begin{aligned}\left\langle \exp\left[it\frac{W_1}{\sqrt{n}}\right]\right\rangle &= \left\langle \sum_{r=0}^{\infty}\frac{(it)^r}{r!}\frac{W_1^r}{n^{r/2}}\right\rangle \\ &= \sum_{r=0}^{\infty}\frac{(it)^r}{r!}\frac{\langle W_1^r\rangle}{n^{r/2}} \\ &\equiv 1 - \frac{t^2}{2n} + \frac{\zeta(t,n)}{n}, \qquad (2.1.4.24)\end{aligned}$$

since $\langle W_1\rangle = 0$, $\langle W_1^2\rangle = 1$, and the quantity denoted by $\zeta(t, n)$ in (2.1.4.24) is given by

$$\zeta(t,n) = \sum_{r=3}^{\infty}\frac{(it)^r}{r!}\frac{\langle W_1^r\rangle}{n^{(r/2)-1}}. \qquad (2.1.4.25)$$

The characteristic function of $\hat{S}_n$ is therefore

$$\langle \exp(it\hat{S}_n)\rangle = \left[1 - \frac{t^2}{2n} + \frac{\zeta(t,n)}{n}\right]^n. \qquad (2.1.4.26)$$

Now, as is seen from (2.1.4.25), for every fixed t the quantity $\zeta(t, n)$ tends to zero as n tends to infinity. The cumulant-generating function of the standardized sum then becomes

$$\log C_n(\hat{S}_n, t) = n\log\left[1 - \frac{1}{n}\left(\frac{t^2}{2} - \zeta(t,n)\right)\right] \qquad (2.1.4.27)$$

and the logarithm on the right-hand side of equation (2.1.4.27) has the form $\log(1 - z)$ with $|z| \to 0$ as $n \to \infty$. We may therefore use the expansion

$$\log(1-z) = -\left(z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots\right),$$

which is valid for $|z| < 1$. We then obtain

$$\begin{aligned}\log C_n(\hat{S}_n, t) &= -n\left[\frac{1}{n}\left(\frac{t^2}{2} - \zeta(t,n)\right) + \frac{1}{2n^2}\left(\frac{t^2}{2} - \zeta(t,n)\right)^2\right. \\ &\quad \left. + \frac{1}{3n^3}\left(\frac{t^2}{2} - \zeta(t,n)\right)^3 + \cdots\right] \\ &= -\frac{t^2}{2} + \zeta(t,n) - \frac{1}{2n}\left(\frac{t^2}{2} - \zeta(t,n)\right)^2 \\ &\quad - \frac{1}{3n^2}\left(\frac{t^2}{2} - \zeta(t,n)\right)^3 - \cdots\end{aligned}$$

and finally, for every fixed t,

$$\lim_{n\to\infty}\log C_n(\hat{S}_n, t) = -\frac{t^2}{2}. \qquad (2.1.4.28)$$

Since the logarithm is a continuous function of t, it follows directly that

$$\lim_{n\to\infty} C_n(\hat{S}_n, t) = \exp\left(-\frac{t^2}{2}\right). \qquad (2.1.4.29)$$

The right-hand side of (2.1.4.29) is just the characteristic function of a standardized normal p.d.f., *i.e.*, a normal p.d.f. with zero mean and unit variance [*cf.* equation (2.1.4.5)]. The asymptotic expression for the p.d.f. of the standardized sum is therefore obtained as

$$p(\hat{S}) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\hat{S}^2}{2}\right),$$

which proves the above version of the central-limit theorem.

Surprisingly, this theorem has a very wide applicability and values of n as low as 30 are often large enough for the theorem to be useful. Situations in which the normal p.d.f. must be modified or replaced by an altogether different one are dealt with in Sections 2.1.7 and 2.1.8 of this chapter.

2.1.4.4. Conditions of validity

The above outline of a proof of the central-limit theorem depended on the existence of moments of all orders. The components of structure factors always possess finite moments of all orders, but the existence of moments beyond the second is not necessary for the validity of the theorem and it can be proved under much less stringent conditions. In fact, if all the random variables in equation (2.1.4.19) have the *same* distribution – as in a homoatomic structure – the only requirement is that the second moments of the distributions should exist [the Lindeberg–Lévy theorem (*e.g.* Cramér, 1951)]. If the distributions are not the same – as in a heteroatomic structure – some further condition is necessary to ensure that no individual random variable dominates the sum. The Liapounoff proof requires the existence of third absolute moments, but this is regarded as aesthetically displeasing; a theorem that ultimately involves only means and variances should require only means and variances in the proof. The Lindeberg–Cramér conditions meet this aesthetic criterion. Roughly, the conditions are that S^2, the variance of the sum, should tend to infinity and σ_j^2/S^2, where σ_j^2 is the variance of the jth random variable, should tend to zero for all j as n tends to infinity. The precise formulation is quoted by Kendall & Stuart (1977, p. 207).

2.1.4.5. Non-independent variables

The central-limit theorem, under certain conditions, remains valid even when the variables summed in equation (2.1.4.19) are not independent. The conditions have been investigated by Bernstein (1922, 1927); roughly they amount to requiring that the

variables should not be too closely correlated. The theorem applies, in particular, when each x_r is related to a finite number, $f(n)$, of its neighbours, when the x's are said to be $f(n)$ dependent. The $f(n)$ dependence seems plausible for crystallographic applications, since the positions of atoms close together in a structure are closely correlated by interatomic forces, whereas those far apart will show little correlation if there is any flexibility in the asymmetric unit when unconstrained. Harker's (1953) idea of 'globs' seems equivalent to $f(n)$ dependence. Long-range stereochemical effects, as in pseudo-graphitic aromatic hydrocarbons, would presumably produce long-range correlations and make $f(n)$ dependence less plausible. If Bernstein's conditions are satisfied, the central-limit theorem would apply, but the actual value of $\langle x^2\rangle - \langle x\rangle^2$ would have to be used for the variance, instead of the sum of the variances of the random variables in (2.1.4.19). Because of the correlations the two values are no longer equal.

French & Wilson (1978) seem to have been the first to appeal explicitly to the central-limit theorem extended to non-independent variables, but many previous workers [for typical references, see Wilson (1981)] tacitly made the replacement – in the X-ray case substituting the local mean intensity for the sum of the squares of the atomic scattering factors.

2.1.5. Ideal probability density distributions

In applications of the central-limit theorem, and its extensions, to intensity statistics the x_j's of equation (2.1.4.19) have the form (atomic scattering factor of the jth atom) times (a trigonometric expression characteristic of the space group and Wyckoff position; also known as the trigonometric structure factor). These trigonometric expressions for all the space groups, and general Wyckoff positions, are given in Tables A1.4.3.1 through A1.4.3.7, and their first few even moments (fixed-index averaging) are given in Table 2.1.7.1. One cannot, of course, conclude that the magnitudes of the structure factor always have a normal distribution – even if the structure is homoatomic; one must look at each problem and see what components of the structure factor can be put in the form (2.1.4.19), deduce the m and σ^2 to be used for each, and combine the components to obtain the asymptotic (large N, not large x) expression for the problem in question. Ordinarily the components are the real and the imaginary parts of the structure factor; the structure factor is purely real only if the structure is centrosymmetric, the space-group origin is chosen at a crystallographic centre and the atoms are nondispersive.

2.1.5.1. Ideal acentric distributions

The ideal acentric distributions are obtained by applying the central-limit theorem to the real and the imaginary parts of the structure factor, as given by equation (2.1.1.1). Consider first a crystal with no rotational symmetry (space group $P1$). The real part, A, of the structure factor is then given by

$$A = \sum_{j=1}^{N} f_j \cos\vartheta_j, \tag{2.1.5.1}$$

where N is the number of atoms in the unit cell and ϑ_j is the phase angle of the jth atom. The central-limit theorem then states that A tends to be normally distributed about its mean value with variance equal to its mean-square deviation from its mean. Under the assumption that the phase angles ϑ_j are uniformly distributed on the 0–2π range, the mean value of each cosine is zero, so that its variance is

$$\sigma^2 = \sum_{j=1}^{N} f_j^2 \langle\cos^2\vartheta_j\rangle. \tag{2.1.5.2}$$

Under the same assumption, the mean value of each $\cos^2\vartheta$ is one-half, so that the variance becomes

$$\sigma^2 = (1/2)\sum_{j=1}^{N} f_j^2 = (1/2)\Sigma, \tag{2.1.5.3}$$

where Σ is the sum of the squares of the atomic scattering factors [*cf.* equation (2.1.2.4)]. The asymptotic form of the distribution of A is therefore given by

$$p(A)\,\mathrm{d}A = (\pi\Sigma)^{-1/2}\exp(-A^2/\Sigma)\,\mathrm{d}A. \tag{2.1.5.4}$$

A similar calculation, with sines instead of cosines, gives an analogous distribution for the imaginary part B, so that the joint probability of the real and imaginary parts of F is

$$p(A, B)\,\mathrm{d}A\,\mathrm{d}B = (\pi\Sigma)^{-1}\exp[-(A^2 + B^2)/\Sigma]\,\mathrm{d}A\,\mathrm{d}B. \tag{2.1.5.5}$$

Ordinarily, however, we are more interested in the distribution of the magnitude, $|F|$, of the structure factor than in the distribution of A and B. Using polar coordinates in equation (2.1.5.5) [$A = |F|\cos\phi$, $B = |F|\sin\phi$] and integrating over the angle ϕ gives

$$p(|F|)\,\mathrm{d}|F| = (2|F|/\Sigma)\exp(-|F|^2/\Sigma)\,\mathrm{d}|F|. \tag{2.1.5.6}$$

It is usually convenient, in structure-factor and intensity statistics, to express the results in terms of the normalized structure factor E and its magnitude $|E|$. If $|F|$ has been put on an absolute scale (see Section 2.2.4.3), we have

$$E = \frac{F}{\sqrt{\Sigma}} \quad \text{and} \quad |E| = \frac{|F|}{\sqrt{\Sigma}}, \tag{2.1.5.7}$$

so that

$$p(|E|)\,\mathrm{d}|E| = 2|E|\exp(-|E|^2)\,\mathrm{d}|E| \tag{2.1.5.8}$$

is the normalized-structure-factor version of (2.1.5.6).

Distributions resulting from noncentrosymmetric crystals are known as *acentric* distributions; those arising from centrosymmetric crystals are known as *centric*. These adjectives are used to describe *distributions*, not crystal symmetry.

2.1.5.2. Ideal centric distributions

When a nondispersive crystal is centrosymmetric, and the space-group origin is chosen at a crystallographic centre of symmetry, the imaginary part B of its structure amplitude is zero. In the simplest case, space group $P\bar{1}$, the contribution of the jth atom plus its centrosymmetric counterpart is $2f_j\cos\vartheta_j$. The calculation of $p(A)$ goes through as before, with allowance for the fact that there are $N/2$ pairs instead of N independent atoms, giving

$$p(A)\,\mathrm{d}A = (2\pi\Sigma)^{-1/2}\exp[-A^2/(2\Sigma)]\,\mathrm{d}A \tag{2.1.5.9}$$

or equivalently

$$p(|F|)\,\mathrm{d}|F| = [2/(\pi\Sigma)]^{1/2}\exp[-|F|^2/(2\Sigma)]\,\mathrm{d}|F| \tag{2.1.5.10}$$

or

$$p(|E|)\,\mathrm{d}|E| = (2/\pi)^{1/2}\exp(-|E|^2/2)\,\mathrm{d}|E|. \tag{2.1.5.11}$$

2.1.5.3. Effect of other symmetry elements on the ideal acentric and centric distributions

Additional crystallographic symmetry elements do not produce any essential alterations in the ideal centric or acentric distribution; their main effect is to replace the parameter Σ by a 'distribution parameter', called S by Wilson (1950) and Rogers

Table 2.1.5.1. *Some properties of gamma and beta distributions*

If $x_1, x_2, \ldots, x_n$ are independent gamma-distributed variables with parameters $p_1, p_2, \ldots, p_n$, their sum is a gamma-distributed variable with $p = p_1 + p_2 + \ldots + p_n$.

If x and y are independent gamma-distributed variables with parameters p and q, then the ratio $u = x/y$ has the distribution $\beta_2(u;\ p, q)$.

With the same notation, the ratio $v = x/(x + y)$ has the distribution $\beta_1(v;\ p, q)$.

Differences and products of gamma-distributed variables do not lead to simple results. For proofs, details and references see Kendall & Stuart (1977).

Name of the distribution, its functional form, mean and variance
Gamma distribution with parameter p: $\gamma_p(x) = [\Gamma(x)]^{-1} x^{p-1} \exp(-x); \quad p \le x \le \infty, \quad p > 0$ mean: $\langle x \rangle = p$; variance: $\langle (x - \langle x \rangle)^2 \rangle = p$.
Beta distribution of first kind with parameters p and q: $\beta_1(x;\ p, q) = \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}; \quad 0 \le x \le \infty, \quad p, q > 0$ mean: $\langle x \rangle = p/(p+q)$; variance: $\langle (x - \langle x \rangle)^2 \rangle = pq/[(p+q)^2(p+q+1)]$.
Beta distribution of second kind with parameters p and q: $\beta_2(x;\ p, q) = \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1+x)^{-p-q}; \quad 0 \le x \le \infty, \quad p, q > 0$ mean: $\langle x \rangle = p/(q-1)$; variance: $\langle (x - \langle x \rangle)^2 \rangle = p(p+q-1)/[(q-1)(q-2)]$.

(1950), in certain groups of reflections. In addition, in noncentrosymmetric space groups, the distribution of certain groups of reflections becomes centric, though the general reflections remain acentric. The changes are summarized in Tables 2.1.3.1 and 2.1.3.2. The values of S are integers for lattice centring, glide planes and those screw axes that produce absences, and approximate integers for rotation axes and mirror planes; the modulations of the average intensity in reciprocal space outlined in Section 2.1.3.2 apply.

It should be noted that if intensities are normalized to the average of the group to which they belong, rather than to the general average, the distributions given in equations (2.1.5.8) and (2.1.5.11) are not affected.

2.1.5.4. Other ideal distributions

The distributions just derived are asymptotic, as they are limiting values for large N. They are the only ideal distributions, in this sense, when there is only strict crystallographic symmetry and no dispersion. However, other ideal (asymptotic) distributions arise when there is noncrystallographic symmetry, or if there is dispersion. The *subcentric* distribution,

$$p(|E|)\,\mathrm{d}|E| = \frac{2|E|}{(1-k^2)^{1/2}} \exp[-|E|^2/(1-k^2)] \times I_0\left(\frac{k|E|^2}{1-k^2}\right) \mathrm{d}|E|, \qquad (2.1.5.12)$$

where $I_0(x)$ is a modified Bessel function of the first kind and k is the ratio of the scattering from the centrosymmetric part to the total scattering, arises when a noncentrosymmetric crystal contains centrosymmetric parts or when dispersion introduces effective noncentrosymmetry into the scattering from a centrosymmetric crystal (Srinivasan & Parthasarathy, 1976, ch. III; Wilson, 1980*a*,*b*; Shmueli & Wilson, 1983). The *bicentric* distribution

$$p(|E|)\,\mathrm{d}|E| = \pi^{-3/2} \exp(-|E|^2/8) K_0(|E|^2/8)\,\mathrm{d}|E| \qquad (2.1.5.13)$$

arises, for example, when the 'asymmetric unit in a centrosymmetric crystal is a centrosymmetric molecule' (Lipson & Woolfson, 1952); $K_0(x)$ is a modified Bessel function of the second kind. There are higher hypercentric, hyperparallel and sesquicentric analogues (Wilson, 1952; Rogers & Wilson, 1953; Wilson, 1956). The ideal subcentric and bicentric distributions are expressed in terms of known functions, but the higher hypercentric and the sesquicentric distributions have so far been studied only through their moments and integral representations. Certain hypersymmetric distributions can be expressed in terms of Meijer's G functions (Wilson, 1987*b*).

2.1.5.5. Relation to distributions of I

When only the intrinsic probability distributions are being considered, it does not greatly matter whether the variable chosen is the intensity of reflection (I), or its positive square root, the modulus of the structure factor ($|F|$), since both are necessarily real and non-negative. In an obvious notation, the relation between the intensity distribution and the structure-factor distribution is

$$p_I(I) = (1/2) I^{-1/2} p_{|F|}(I^{1/2}) \qquad (2.1.5.14)$$

or

$$p_{|F|}(|F|) = 2|F| p_I(|F|^2). \qquad (2.1.5.15)$$

Statistical fluctuations in counting rates, however, introduce a small but finite probability of negative observed intensities (Wilson, 1978*a*, 1980*a*) and thus of imaginary structure factors. This practical complication is treated in *IT* C (2004, Parts 7 and 8).

Both the ideal centric and acentric distributions are simple members of the family of gamma distributions, defined by

$$\gamma_n(x)\,\mathrm{d}x = [\Gamma(n)]^{-1} x^{n-1} \exp(-x)\,\mathrm{d}x, \qquad (2.1.5.16)$$

where n is a parameter, not necessarily integral, and $\Gamma(n)$ is the gamma function. Thus the ideal acentric intensity distribution is

$$p(I)\,\mathrm{d}I = \exp(-I/\Sigma)\,\mathrm{d}(I/\Sigma) \qquad (2.1.5.17)$$

$$= \gamma_1(I/\Sigma)\,\mathrm{d}(I/\Sigma) \qquad (2.1.5.18)$$

and the ideal centric intensity distribution is

$$p(I)\,\mathrm{d}I = (2\Sigma/\pi)^{1/2} \exp[-I/(2\Sigma)]\,\mathrm{d}[I/(2\Sigma)] \qquad (2.1.5.19)$$

$$= \gamma_{1/2}[I/(2\Sigma)]\,\mathrm{d}[I/(2\Sigma)]. \qquad (2.1.5.20)$$

The properties of gamma distributions and of the related beta distributions, summarized in Table 2.1.5.1, are used in Section 2.1.6 to derive the probability density functions of sums and of ratios of intensities drawn from one of the ideal distributions.

2.1.5.6. Cumulative distribution functions

The integral of the probability density function $f(x)$ from the lower end of its range up to an arbitrary value x is called the cumulative probability distribution, or simply the distribution function, $F(x)$, of x. It can always be written

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u; \qquad (2.1.5.21)$$

if the lower end of its range is not actually $-\infty$ one takes $f(x)$ as identically zero between $-\infty$ and the lower end of its range. For the distribution of A [equation (2.1.5.4) or (2.1.5.9)] the lower

limit is in fact $-\infty$; for the distribution of $|F|$, $|E|$, I and I/Σ the lower end of the range is zero. In such cases, equation (2.1.5.21) becomes

$$F(x) = \int_0^x f(x)\,\mathrm{d}x. \qquad (2.1.5.22)$$

In crystallographic applications the cumulative distribution is usually denoted by $N(x)$, rather than by the capital letter corresponding to the probability density function designation. The cumulative forms of the ideal acentric and centric distributions (Howells *et al.*, 1950) have found many applications. For the acentric distribution of $|E|$ [equation (2.1.5.8)] the integration is readily carried out:

$$N(|E|) = 2\int_0^{|E|} y\exp(-y^2)\,\mathrm{d}y = 1 - \exp(-|E|^2). \qquad (2.1.5.23)$$

The integral for the centric distribution of $|E|$ [equation (2.1.5.11)] cannot be expressed in terms of elementary functions, but the integral required has so many important applications in statistics that it has been given a special name and symbol, the error function $\mathrm{erf}(x)$, defined by

$$\mathrm{erf}(x) = (2/\pi^{1/2})\int_0^x \exp(-t^2)\,\mathrm{d}t. \qquad (2.1.5.24)$$

For the centric distribution, then

$$N(|E|) = (2/\pi)^{1/2}\int_0^{|E|} y\exp(-y^2/2)\,\mathrm{d}y \qquad (2.1.5.25)$$

$$= \mathrm{erf}(|E|/2^{1/2}). \qquad (2.1.5.26)$$

The error function is extensively tabulated [see *e.g.* Abramowitz & Stegun (1972), pp. 310–311, and a closely related function on pp. 966–973].

The foregoing and following structure-factor and intensity statistics, presented in this chapter, are centred around the unambiguous determination of crystal symmetry. These statistics, especially the ideal ones, are routinely applied to X-ray diffraction data for completing the symmetry information where needed. On the other hand, the symmetry of very fine grained, imperfectly crystalline material or individual nanostructures can be experimentally determined by the convergent-beam electron diffraction technique using an electron microscope with a nanoscale probe, as described and illustrated in Section 2.5.3 of this volume.

2.1.6. Distributions of sums, averages and ratios

2.1.6.1. Distributions of sums and averages

In Section 2.1.2.1, it was shown that the average intensity of a sufficient number of reflections is Σ [equation (2.1.2.4)]. When the number of reflections is not 'sufficient', their mean value will show statistical fluctuations about Σ; such statistical fluctuations are in addition to any systematic variation resulting from non-independence of atomic positions, as discussed in Sections 2.1.2.1–2.1.2.3. We thus need to consider the probability density functions of sums like

$$J_n = \sum_{i=1}^{n} G_i, \qquad (2.1.6.1)$$

and averages like

$$Y = J_n/n, \qquad (2.1.6.2)$$

where G_i is the intensity of the ith reflection. The probability density distributions are easily obtained from a property of gamma distributions: If $x_1, x_2, \ldots, x_n$ are independent gamma-distributed variables with parameters $p_1, p_2, \ldots, p_n$, their sum is a gamma-distributed variable with parameter p equal to the sum of the parameters. The sum of n intensities drawn from an acentric distribution thus has the distribution

$$p(J_n)\,\mathrm{d}J_n = \gamma_n(J_n/\Sigma)\,\mathrm{d}(J_n/\Sigma); \qquad (2.1.6.3)$$

the parameters of the variables added are all equal to unity, so that their sum is p. Similarly, the sum of n intensities drawn from a centric distribution has the distribution

$$p(J_n)\,\mathrm{d}J_n = \gamma_{n/2}[J_n/(2\Sigma)]\,\mathrm{d}[J_n/(2\Sigma)]; \qquad (2.1.6.4)$$

each parameter has the value of one-half. The corresponding distributions of the averages of n intensities are then

$$p(Y)\,\mathrm{d}Y = \gamma_n(nY/\Sigma)\,\mathrm{d}(nY/\Sigma) \qquad (2.1.6.5)$$

for the acentric case, and

$$p(Y)\,\mathrm{d}Y = \gamma_{n/2}[nY/(2\Sigma)]\,\mathrm{d}[nY/(2\Sigma)] \qquad (2.1.6.6)$$

for the centric. In both cases the expected value of Y is Σ and the variances are Σ^2/n and $2\Sigma^2/n$, respectively, just as would be expected.

2.1.6.2. Distribution of ratios

Ratios like

$$S_{n,m} = J_n/K_m, \qquad (2.1.6.7)$$

where J_n is given by equation (2.1.6.1),

$$K_m = \sum_{j=1}^{m} H_j, \qquad (2.1.6.8)$$

and the H_j's are the intensities of a set of reflections (which may or may not overlap with those included in J_n), are used in correlating intensities measured under different conditions. They arise in correlating reflections on different layer lines from the same or different specimens, in correlating the same reflections from different crystals, in normalizing intensities to the local average or to Σ, and in certain systematic trial-and-error methods of structure determination (see Rabinovich & Shakked, 1984, and references therein). There are three main cases:

(i) G_i and H_i refer to the *same* reflection; for example, they might be the observed and calculated quantities for the *hkl* reflection measured under different conditions or for different crystals of the same substance; or

(ii) G_i and H_i are *unrelated*; for example, the observed and calculated values for the *hkl* reflection for a completely wrong trial structure, of values for entirely different reflections, as in reducing photographic measurements on different layer lines to the same scale; or

(iii) the G_i's are a subset of the H_i's, so that $G_i = H_i$ for $i < n$ and $m > n$.

Aside from the scale factor, in case (i) G_i and H_i will differ chiefly through relatively small statistical fluctuations and uncorrected systematic errors, whereas in case (ii) the differences will be relatively large because of the inherent differences in the intensities. Here we are concerned only with cases (ii) and (iii); the practical problems of case (i) are postponed to *IT* C (2004), Chapter 7.5.

There is little in the crystallographic literature concerning the probability distribution of sums like (2.1.6.1) or ratios like (2.1.6.7); certain results are reviewed by Srinivasan & Parthasarathy (1976, ch. 5), but with a bias toward partially related structures that makes it difficult to apply them to the immediate problem.

In case (ii) (G_i and H_i independent), acentric distribution, Table 2.1.5.1 gives the distribution of the ratio

$$u = nY/(mZ) \tag{2.1.6.9}$$

$$p(u)\,\mathrm{d}u = \beta_2[nY/(mZ);\, n, m]\,\mathrm{d}[nY/(mZ)], \tag{2.1.6.10}$$

where β_2 is a beta distribution of the second kind, Y is given by equation (2.1.6.2) and Z by

$$Z = K_m/m, \tag{2.1.6.11}$$

where n is the number of intensities included in the numerator and m is the number in the denominator. The expected value of Y/Z is then

$$\langle Y/Z\rangle = \frac{m}{m-1} = 1 + \frac{1}{m} + \ldots \tag{2.1.6.12}$$

with variance

$$\sigma^2 = \frac{(n+m-1)m^2}{(m-1)^2(m-2)n}. \tag{2.1.6.13}$$

One sees that Y/Z is a biased estimate of the scaling factor between two sets of intensities and the bias, of the order of m^{-1}, depends only on the number of intensities averaged in the denominator. This may seem odd at first sight, but it becomes plausible when one remembers that the mean of a quantity is an unbiased estimator of itself, but the reciprocal of a mean is not an unbiased estimator of the mean of a reciprocal. The mean exists only if $m > 1$ and the variance only for $m > 2$.

In the centric case, the expression for the distribution of the ratio of the two means Y and Z becomes

$$p(u)\,\mathrm{d}u = \beta_2[nY/(mZ);\ n/2, m/2]\,\mathrm{d}[nY/(mZ)] \tag{2.1.6.14}$$

with the expected value of Y/Z equal to

$$\langle Y/Z\rangle = \frac{m}{m-2} = 1 + \frac{2}{m} + \ldots \tag{2.1.6.15}$$

and with its variance equal to

$$\sigma^2 = \frac{2(n+m-2)m^2}{(m-2)^2(m-4)n}. \tag{2.1.6.16}$$

For the same number of reflections, the bias in $\langle Y/Z\rangle$ and the variance for the centric distribution are considerably larger than for the acentric. For both distributions the variance of the scaling factor approaches zero when n and m become large. The variances are large for m small, in fact 'infinite' if the number of terms averaged in the denominator is sufficiently small. These biases are readily removed by multiplying Y/Z by $(m-1)/m$ or $(m-2)/m$. Many methods of estimating scaling factors – perhaps most – also introduce bias (Wilson, 1975; Lomer & Wilson, 1975; Wilson, 1976, 1978*c*) that is not so easily removed. Wilson (1986*a*) has given reasons for supposing that the bias of the ratio (2.1.6.7) approximates to

$$1 + \frac{\sigma^2(I)}{m\langle I\rangle^2}, \tag{2.1.6.17}$$

whatever the intensity distribution. Equations (2.1.6.12) and (2.1.6.15) are consistent with this.

2.1.6.3. Intensities scaled to the local average

When the G_i's are a subset of the H_i's, the beta distributions of the second kind are replaced by beta distributions of the first kind, with means and variances readily found from Table 2.1.5.1. The distribution of such a ratio is chiefly of interest when Y relates to a single reflection and Z relates to a group of m intensities including Y. This corresponds to normalizing intensities to the local average. Its distribution is

$$p(I/\langle I\rangle)\,\mathrm{d}(I/\langle I\rangle) = \beta_1(I/n\langle I\rangle;\ 1, n-1)\,\mathrm{d}(I/n\langle I\rangle) \tag{2.1.6.18}$$

in the acentric case, with an expected value of $I/\langle I\rangle$ of unity; there is no bias, as is obvious *a priori*. The variance of $I/\langle I\rangle$ is

$$\sigma^2 = \frac{n-1}{n+1}, \tag{2.1.6.19}$$

which is less than the variance of the intensities normalized to an 'infinite' population by a fraction of the order of $2/n$. Unlike the variance of the scaling factor, the variance of the normalized intensity approaches unity as n becomes large. For intensities having a centric distribution, the distribution normalized to the local average is given by

$$p(I/\langle I\rangle)\,\mathrm{d}(I/\langle I\rangle) = \beta_1[I/n\langle I\rangle;\ 1/2, (n-1)/2]\,\mathrm{d}(I/n\langle I\rangle), \tag{2.1.6.20}$$

with an expected value of $I/\langle I\rangle$ of unity and with variance

$$\sigma^2 = \frac{2(n-1)}{n+2}, \tag{2.1.6.21}$$

less than that for an 'infinite' population by a fraction of about $3/n$.

Similar considerations apply to intensities normalized to Σ in the usual way, since they are equal to those normalized to $\langle I\rangle$ multiplied by $\langle I\rangle/\Sigma$.

2.1.6.4. The use of normal approximations

Since J_n and K_m [equations (2.1.6.1) and (2.1.6.8)] are sums of identically distributed variables conforming to the conditions of the central-limit theorem, it is tempting to approximate their distributions by normal distributions with the correct mean and variance. This would be reasonably satisfactory for the distributions of J_n and K_m themselves for quite small values of n and m, but unsatisfactory for the distribution of their ratio for any values of n and m, even large. The ratio of two variables with normal distributions is notorious for its rather indeterminate mean and 'infinite' variance, resulting from the 'tail' of the denominator distributions extending through zero to negative values. The leading terms of the ratio distribution are given by Kendall & Stuart (1977, p. 288).

2.1.7. Non-ideal distributions: the correction-factor approach

2.1.7.1. Introduction

The probability density functions (p.d.f.'s) of the magnitude of the structure factor, presented in Section 2.1.5, are based on the central-limit theorem discussed above. In particular, the centric and acentric p.d.f.'s given by equations (2.1.5.11) and (2.1.5.8), respectively, are expected to account for the statistical properties of diffraction patterns obtained from crystals consisting of nearly equal atoms, which obey the fundamental assumptions of uniformity and independence of the atomic contributions and are not affected by noncrystallographic symmetry and dispersion. It is also assumed there that the number of atoms in the asymmetric unit is large. Distributions of structure-factor magnitudes which are based on the central-limit theorem, and thus obey the above assumptions, have been termed 'ideal', and the subjects of the following sections are those distributions for which some of the above assumptions/restrictions are not fulfilled; the latter distributions will be called 'non-ideal'.

We recall that the assumption of uniformity consists of the requirement that the fractional part of the scalar product $hx + ky + lz$ be uniformly distributed over the [0, 1] interval, which holds well if x, y, z are rationally independent (Hauptman & Karle, 1953), and permits one to regard the atomic contribu-

tion to the structure factor as a random variable. This is of course a necessary requirement for any statistical treatment. If, however, the atomic composition of the asymmetric unit is widely heterogeneous, the structure factor is then a sum of unequally distributed random variables and the Lindeberg–Lévy version of the central-limit theorem (*cf.* Section 2.1.4.4) cannot be expected to apply. Other versions of this theorem might still predict a normal p.d.f. of the sum, but at the expense of a correspondingly large number of terms/atoms. It is well known that atomic heterogeneity gives rise to severe deviations from ideal behaviour (*e.g.* Howells *et al.*, 1950) and one of the aims of crystallographic statistics has been the introduction of a correct dependence on the atomic composition into the non-ideal p.d.f.'s [for a review of the early work on non-ideal distributions see Srinivasan & Parthasarathy (1976)]. A somewhat less well known fact is that the dependence of the p.d.f.'s of $|E|$ on space-group symmetry becomes more conspicuous as the composition becomes more heterogeneous (*e.g.* Shmueli, 1979; Shmueli & Wilson, 1981). Hence both the composition and the symmetry dependence of the intensity statistics are of interest. Other problems, which likewise give rise to non-ideal p.d.f.'s, are the presence of heavy atoms in (variable) special positions, heterogeneous structures with complete or partial noncrystallographic symmetry, and the presence of outstandingly heavy dispersive scatterers.

The need for theoretical representations of non-ideal p.d.f.'s is exemplified in Fig. 2.1.7.1, which shows the ideal centric and acentric p.d.f.'s together with a frequency histogram of $|E|$ values, recalculated for a centrosymmetric structure containing a platinum atom in the asymmetric unit of $P\bar{1}$ (Faggiani *et al.*, 1980). Clearly, the deviation from the Gaussian p.d.f., predicted by the central-limit theorem, is here very large and a comparison with the possible ideal distributions can (in this case) lead to wrong conclusions.

Two general approaches have so far been employed in derivations of non-ideal p.d.f.'s which account for the above-mentioned problems: the correction-factor approach, to be dealt with in the following sections, and the more recently introduced Fourier method, to which Section 2.1.8 is dedicated. In what follows, we introduce briefly the mathematical background of the correction-factor approach, apply this formalism to centric and acentric non-ideal p.d.f.'s, and present the numerical values of the moments of the trigonometric structure factor which permit an approximate evaluation of such p.d.f.'s for all the three-dimensional space groups.

2.1.7.2. Mathematical background

Suppose that $p(x)$ is a p.d.f. which accurately describes the experimental distribution of the random variable x, where x is related to a sum of random variables and can be assumed to obey (to some approximation) an ideal p.d.f., say $p^{(0)}(x)$, based on the central-limit theorem. In the correction-factor approach we seek to represent $p(x)$ as

$$p(x) = p^{(0)}(x) \sum_k d_k f_k(x), \qquad (2.1.7.1)$$

where d_k are coefficients which depend on the cause of the deviation of $p(x)$ from the central-limit theorem approximation and $f_k(x)$ are suitably chosen functions of x. A choice of the set $\{f_k\}$ is deemed suitable, if only from a practical point of view, if it allows the convenient introduction of the cause of the above deviation of $p(x)$ into the expansion coefficients d_k. This requirement is satisfied – also from a theoretical point of view – by taking $f_k(x)$ as a set of polynomials which are orthogonal with respect to the ideal p.d.f., taken as their weight function (*e.g.* Cramér, 1951). That is, the functions $f_k(x)$ so chosen have to obey the relationship

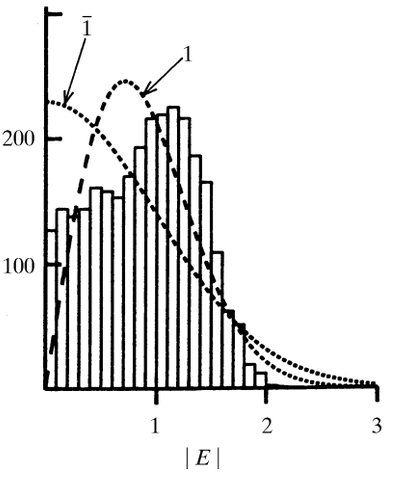


Fig. 2.1.7.1. Atomic heterogeneity and intensity statistics. The histogram appearing in this figure was constructed from $|E|$ values which were recalculated from atomic parameters published for the centrosymmetric structure of $C_6H_{18}Cl_2N_4O_4Pt$ (Faggiani *et al.*, 1980). The space group of the crystal is $P\bar{1}$, $Z = 2$, *i.e.* all the atoms are located in general positions. The figure shows a comparison of the recalculated distribution of $|E|$ with the ideal centric [equation (2.1.5.11)] and acentric [equation (2.1.5.8)] p.d.f.'s, denoted by $\bar{1}$ and 1, respectively.

$$\int_a^b f_k(x) f_m(x) p^{(0)}(x)\,\mathrm{d}x = \delta_{km} = \begin{cases} 1, & \text{if} \quad k = m \\ 0, & \text{if} \quad k \neq m \end{cases}, \qquad (2.1.7.2)$$

where $[a, b]$ is the range of existence of all the functions involved. It can be readily shown that the coefficients d_k are given by

$$d_k = \int_a^b f_k(x) p(x)\,\mathrm{d}x = \langle f_k(x) \rangle = \sum_{n=0}^{k} c_n^{(k)} \langle x^n \rangle, \qquad (2.1.7.3)$$

where the brackets $\langle\ \rangle$ in equation (2.1.7.3) denote averaging with respect to the *unknown* p.d.f. $p(x)$ and $c_n^{(k)}$ is the coefficient of the nth power of x in the polynomial $f_k(x)$. The coefficients d_k are thus directly related to the moments of the non-ideal distribution and the coefficients of the powers of x in the orthogonal polynomials. The latter coefficients can be obtained by the Gram–Schmidt procedure (*e.g.* Spiegel, 1974), or by direct use of the Szegö determinants (*e.g.* Cramér, 1951), for any weight function that has finite moments. However, the feasibility of the present approach depends on our ability to obtain the moments $\langle x^n \rangle$ *without* the knowledge of the non-ideal p.d.f., $p(x)$.

2.1.7.3. Application to centric and acentric distributions

We shall summarize here the non-ideal centric and acentric distributions of the magnitude of the normalized structure factor E (*e.g.* Shmueli & Wilson, 1981; Shmueli, 1982). We assume that (i) all the atoms are located in general positions and have rationally independent coordinates, (ii) all the scatterers are dispersionless, and (iii) there is no noncrystallographic symmetry. Arbitrary atomic composition and space-group symmetry are admitted. The appropriate weight functions and the corresponding orthogonal polynomials are

$p^{(0)}(\|E\|)$	$f_k(x)$	Non-ideal distribution
$(2/\pi)^{1/2} \exp(-\|E\|^2/2)$	$He_{2k}(\|E\|)/[(2k)!]^{1/2}$	Centric
$2\|E\| \exp(-\|E\|^2)$	$L_k(\|E\|^2)$	Acentric

(2.1.7.4)

where He_k and L_k are Hermite and Laguerre polynomials, respectively, as defined, for example, by Abramowitz & Stegun (1972). Equations (2.1.7.2), (2.1.7.3) and (2.1.7.4) suffice for the general formulation of the above non-ideal p.d.f.'s of $|E|$. Their full derivation entails (i) the expression of a sufficient number of moments of $|E|$ in terms of absolute moments of the trigonometric structure factor (*e.g.* Shmueli & Wilson, 1981; Shmueli, 1982) and (ii) calculation of the latter moments for the various symmetries (Wilson, 1978*b*; Shmueli & Kaldor, 1981, 1983). The notation below is similar to that employed by Shmueli (1982).

These non-ideal p.d.f.'s of $|E|$, for which the first five expansion terms are available, are given by

$$p_{\rm c}(|E|) = p_{\rm c}^{(0)}(|E|)\left[1 + \sum_{k=2}^{\infty} \frac{A_{2k}}{(2k)!} He_{2k}(|E|)\right] \qquad (2.1.7.5)$$

and

$$p_{\rm a}(|E|) = p_{\rm a}^{(0)}(|E|)\left[1 + \sum_{k=2}^{\infty} \frac{(-1)^k B_{2k}}{k!} L_k(|E|^2)\right] \qquad (2.1.7.6)$$

for centrosymmetric and noncentrosymmetric space groups, respectively, where $p_{\rm c}^{(0)}(|E|)$ and $p_{\rm a}^{(0)}(|E|)$ are the ideal centric and acentric p.d.f.'s [see (2.1.7.4)] and the unified form of the coefficients A_{2k} and B_{2k}, for k = 2, 3, 4 and 5, is

$$\begin{aligned}
&A_4 \quad \text{or} \quad B_4 = a_4 Q_4 \\
&A_6 \quad \text{or} \quad B_6 = a_6 Q_6 \\
&A_8 \quad \text{or} \quad B_8 = a_8 Q_8 + U(a_4^2 Q_4^2 - \gamma_4^2) \\
&A_{10} \quad \text{or} \quad B_{10} = a_{10} Q_{10} + V(a_4 a_6 Q_4 Q_6 - \gamma_4 \gamma_6 Q_{10}) \\
&\qquad\qquad + W\gamma_4^2 Q_{10}
\end{aligned} \qquad (2.1.7.7)$$

(Shmueli, 1982), where U = 35 or 18, V = 210 or 100 and W = 3150 or 900 according as A_{2k} or B_{2k} is required, respectively, and the other quantities in equation (2.1.7.7) are given below. The *composition-dependent* terms in equations (2.1.7.7) are

$$Q_{2k} = \frac{\sum_{j=1}^{m} f_j^{2k}}{\left(\sum_{n=1}^{m} f_n^2\right)^k}, \qquad (2.1.7.8)$$

where m is the number of atoms in the asymmetric unit, f_j, $j = 1, \ldots, m$ are their scattering factors, and the *symmetry* dependence is expressed by the coefficients a_{2k} in equation (2.1.7.7), as follows:

$$a_{2k} = (-1)^{k-1}(k-1)!\alpha_{k0} + \sum_{p=2}^{k}(-1)^{k-p}(k-p)!\alpha_{kp}\gamma_{2p}, \quad (2.1.7.9)$$

where

$$\alpha_{kp} = \binom{k}{p}\frac{(2k-1)!!}{(2p-1)!!} \quad \text{or} \quad \binom{k}{p}\frac{k!}{p!} \qquad (2.1.7.10)$$

according as the space group is centrosymmetric or noncentrosymmetric, respectively, and γ_{2p} in equation (2.1.7.9) is given by

$$\gamma_{2p} = \frac{\langle |T|^{2p}\rangle}{\langle |T|^2\rangle^p}, \qquad (2.1.7.11)$$

where $\langle |T|^k\rangle$ is the kth absolute moment of the trigonometric structure factor

$$T(\mathbf{h}) = \sum_{s=1}^{g} \exp[2\pi i \mathbf{h}^T(\mathbf{P}_s\mathbf{r} + \mathbf{t}_s)] \equiv \xi(\mathbf{h}) + i\eta(\mathbf{h}). \qquad (2.1.7.12)$$

In equation (2.1.7.12), g is the number of general equivalent positions listed in *IT* A (2005) for the space group in question, times the multiplicity of the Bravais lattice, $(\mathbf{P}_s, \mathbf{t}_s)$ is the sth space-group operator and $\mathbf{r}$ is an atomic position vector.

The cumulative distribution functions, obtained by integrating equations (2.1.7.5) and (2.1.7.6), are given by

$$N_{\rm c}(|E|) = \mathrm{erf}\left(\frac{|E|}{\sqrt{2}}\right) - \frac{2}{\sqrt{\pi}}\exp\left(-\frac{|E|^2}{2}\right) \times \left[\sum_{k=2}^{\infty} \frac{A_{2k}}{(2k)!} He_{2k-1}(|E|)\right] \qquad (2.1.7.13)$$

and

$$N_{\rm a}(|E|) = 1 - \exp(-|E|^2) + \exp(-|E|^2) \times \left\{\sum_{k=2}^{\infty} \frac{(-1)^k B_{2k}}{k!}[L_{k-1}(|E|^2) - L_k(|E|^2)]\right\} \qquad (2.1.7.14)$$

for centrosymmetric and noncentrosymmetric space groups, respectively, where the coefficients are defined in equations (2.1.7.7)–(2.1.7.12). Note that the first term on the right-hand side of equation (2.1.7.13) and the first two terms on the right-hand side of equation (2.1.7.14) are just the cumulative distributions derived from the ideal centric and acentric p.d.f.'s in Section 2.1.5.6.

The moments $\langle |T|^{2k}\rangle$ were compiled for all the space groups by Wilson (1978*b*) for $k = 1$ and 2, and by Shmueli & Kaldor (1981, 1983) for k = 1, 2, 3 and 4. These results are presented in Table 2.1.7.1. Closed expressions for the normalized moments γ_{2p} were obtained by Shmueli (1982) for the triclinic, monoclinic and orthorhombic space groups except *Fdd*2 and *Fddd* (see Table 2.1.7.2). The composition-dependent terms, Q_{2k}, are most conveniently computed as weighted averages over the ranges of $(\sin\theta)/\lambda$ which were used in the construction of the Wilson plot for the computation of the $|E|$ values.

2.1.7.4. Fourier versus Hermite approximations

As noted in Section 2.1.8.7 below, the Fourier representation of the probability distribution of $|F|$ is usually much better than the particular orthogonal-function representation discussed in Section 2.1.7.3. Many, perhaps most, non-ideal centric distributions look like slight distortions of the ideal (Gaussian) distribution and have no resemblance to a cosine function. The empirical observation thus seems paradoxical. The probable explanation has been pointed out by Wilson (1986*b*). A truncated Fourier series is a best approximation, in the least-squares sense, to the function represented. The particular orthogonal-function approach used in equation (2.1.7.5), on the other hand, is not a least-squares approximation to $p_{\rm c}(|E|)$, but is a least-squares approximation to

$$p_{\rm c}(|E|)\exp(|E|^2/4). \qquad (2.1.7.15)$$

The usual expansions (often known as Gram–Charlier or Edgeworth) thus give great weight to fitting the distribution of the (compararively few) strong reflections, at the expense of a poor fit for the (much more numerous) weak-to-medium ones. Presumably, a similar situation exists for the representation of acentric distributions, but this has not been investigated in detail. Since the centric distributions $p_{\rm c}(|E|)$ often look nearly Gaussian, one is led to ask if there is an expansion in orthogonal functions that (i) has the leading term $p_{\rm c}(|E|)$ and (ii) is a least-squares (as well as an orthogonal-function)[2] fit to $p_{\rm c}(|E|)$. One does exist, based on the orthogonal functions

[2] The condition for this desirable property seems to be that the weight function in equation (2.1.7.2) should be unity.

Table 2.1.7.1. *Some even absolute moments of the trigonometric structure factor*

The symbols p, q, r and s denote the second, fourth, sixth and eighth absolute moments of the trigonometric structure factor T [equation (2.1.7.12)], respectively, and the columns of the table contain (for some conciseness) p, q, r/p and s/p^2. The numbers in parentheses, appearing beside some space-group entries, refer to *hkl* subsets which are defined in the note at the end of the table. These subset references are identical with those given by Shmueli & Kaldor (1981, 1983). The symbols q, r and s are also equivalent to $\gamma_4 P^2$, $\gamma_6 P^3$ and $\gamma_8 P^4$, respectively, where γ_{2n} are the normalized absolute moments given by equation (2.1.7.11).

Space groups(s)	p	q	r/p	s/p^2
Point group: 1				
$P1$	1	1	1	1
Point group: $\bar{1}$				
$P\bar{1}$	2	6	10	$17\frac{1}{2}$
Point groups: 2, *m*				
All P	2	6	10	$17\frac{1}{2}$
All C	4	48	160	560
Point group: 2/*m*				
All P	4	36	100	$306\frac{1}{4}$
All C	8	288	1600	9800
Point group: 222				
All P	4	28	64	$169\frac{3}{4}$
All C and I	8	224	1024	5432
$F222$	16	1792	16384	173824
Point group: *mm*2				
All P	4	36	100	$306\frac{1}{4}$
All A, C and I	8	288	1600	9800
$Fmm2$	16	2304	25600	313600
$Fdd2$ (1)	16	2304	25600	313600
$Fdd2$ (2)	16	1280	7168	43264
Point group: *mmm*				
All P	8	216	1000	$5359\frac{3}{8}$
All C and I	16	1728	16000	171500
$Fmmm$	32	13824	256000	5488000
$Fddd$ (1)	32	13824	256000	5488000
$Fddd$ (2)	32	7680	71680	757120
Point group: 4				
$P4$, $P4_2$	4	36	100	$306\frac{1}{4}$
$P4_1$† (3)	4	36	100	$306\frac{1}{4}$
$P4_1$† (4)	4	20	28	$42\frac{1}{4}$
$I4$	8	288	1600	9800
$I4_1$ (5)	8	288	1600	9800
$I4_1$ (6)	8	160	448	1352
Point group: $\bar{4}$				
$P\bar{4}$	4	28	64	$169\frac{3}{4}$
$I\bar{4}$	8	224	1024	5432
Point group: 4/*m*				
All P	8	216	1000	$5359\frac{3}{8}$
$I4/m$	16	1728	16000	171500
$I4_1/a$ (7)	16	1728	16000	171500
$I4_1/a$ (8)	16	960	4480	23660
Point group: 422				
$P422$, $P42_12$, $P4_222$, $P4_22_12$	8	136	424	$1682\frac{1}{8}$
$P4_122$,† $P4_12_12$† (3)	8	136	424	$1682\frac{1}{8}$
$P4_122$,† $P4_12_12$† (4)	8	104	208	$470\frac{1}{8}$
$I422$	16	1088	6784	53828
$I4_122$ (7)	16	1088	6784	53828
$I4_122$ (8)	16	832	3328	15044
Point group: 4*mm*				
All P	8	168	640	$2970\frac{5}{8}$
$I4mm$, $I4cm$	16	1344	10240	95060
$I4_1md$, $I4_1cd$ (7)	16	1344	10240	95060
$I4_1md$, $I4_1cd$ (8)	16	832	3328	15188
Point groups: $\bar{4}$2*m*, $\bar{4}$*m*2				
All P	8	136	424	$1682\frac{1}{8}$
$I\bar{4}m2$, $I\bar{4}2m$, $I\bar{4}c2$	16	1088	6784	53828
$I\bar{4}2d$ (5)	16	1088	6784	53828
$I\bar{4}2d$ (6)	16	832	3328	15044
Point group: 4/*mmm*				
All P	16	1008	6400	$51985\frac{15}{16}$
$I4/mmm$, $I4/mcm$	32	8064	102400	1663550
$I4_1/amd$, $I4_1/acd$ (5)	32	8064	102400	1663550
$I4_1/amd$, $I4_1/acd$ (6)	32	4992	33280	265790
Point group: 3				
All P and R	3	15	31	71
Point group: $\bar{3}$				
All P and R	6	90	310	$1242\frac{1}{2}$
Point group: 32				
All P and R	6	66	166	$508\frac{1}{2}$
Point group: 3*m*				
$P3m1$, $P31m$, $R3m$	6	66	178	$604\frac{1}{2}$
$P3c1$, $P31c$ (3), $R3c$ (1)	6	66	178	$604\frac{1}{2}$
$P3c1$, $P31c$ (4), $R3c$ (2)	6	66	154	$412\frac{1}{2}$
Point group: $\bar{3}$*m*				
$P\bar{3}1m$, $P\bar{3}m1$, $R\bar{3}m$	12	396	1780	$10578\frac{3}{4}$
$P\bar{3}1c$, $P\bar{3}c1$ (3), $R\bar{3}c$ (1)	12	396	1780	$10578\frac{3}{4}$
$P\bar{3}1c$, $P\bar{3}c1$ (4), $R\bar{3}c$ (2)	12	396	1540	$7218\frac{3}{4}$
Point group: 6				
$P6$	6	90	340	$1522\frac{1}{2}$
$P6_1$† (9)	6	90	340	$1522\frac{1}{2}$
$P6_1$† (10)	6	54	91	$161\frac{1}{2}$
$P6_1$† (11)	6	54	97	$193\frac{1}{2}$
$P6_1$† (12)	6	90	280	$962\frac{1}{2}$
$P6_2$† (13)	6	90	340	$1522\frac{1}{2}$
$P6_2$† (14)	6	54	97	$193\frac{1}{2}$
$P6_3$ (3)	6	90	340	$1522\frac{1}{2}$
$P6_3$ (4)	6	90	280	$962\frac{1}{2}$
Point group: $\bar{6}$				
$P\bar{6}$	6	90	310	$1242\frac{1}{2}$
Point group: 6/*m*				
$P6/m$	12	540	3400	$26643\frac{3}{4}$
$P6_3/m$ (3)	12	540	3400	$26643\frac{3}{4}$
$P6_3/m$ (4)	12	540	2800	$16843\frac{3}{4}$
Point group: 622				
$P622$	12	324	1150	$5506\frac{1}{4}$
$P6_122$† (9)	12	324	1150	$5506\frac{1}{4}$
$P6_122$† (10)	12	252	577	$1537\frac{3}{4}$
$P6_122$† (11)	12	252	583	$1601\frac{3}{4}$
$P6_122$† (12)	12	324	1090	$4746\frac{1}{4}$
$P6_222$† (13)	12	324	1150	$5506\frac{1}{4}$
$P6_222$† (14)	12	252	583	$1601\frac{3}{4}$
$P6_322$ (3)	12	324	1150	$5506\frac{1}{4}$
$P6_322$ (4)	12	324	1090	$4746\frac{1}{4}$
Point group: 6*mm*				
$P6mm$	12	396	1930	$12818\frac{3}{4}$
$P6cc$ (3)	12	396	1930	$12818\frac{3}{4}$
$P6cc$ (4)	12	396	1450	$6098\frac{3}{4}$
$P6_3cm$, $P6_3mc$ (3)	12	396	1930	$12818\frac{3}{4}$
$P6_3cm$, $P6_3mc$ (4)	12	396	1630	$8338\frac{3}{4}$
Point groups: $\bar{6}$*m*2, $\bar{6}$2*m*				
$P\bar{6}m2$, $P\bar{6}2m$	12	396	1780	$10578\frac{3}{4}$
$P\bar{6}c2$, $P\bar{6}2c$ (3)	12	396	1780	$10578\frac{3}{4}$

Table 2.1.7.1 (*cont.*)

Space groups(s)	p	q	r/p	s/p^2
$P\bar{6}c2$, $P\bar{6}2c$ (4)	12	396	1540	$7218\frac{3}{4}$
Point group: 6/*mmm*				
$P6/mmm$	24	2376	19300	$224328\frac{1}{8}$
$P6/mcc$ (3)	24	2376	19300	$224328\frac{1}{8}$
$P6/mcc$ (4)	24	2376	14500	$106728\frac{1}{8}$
$P6/mcm$, $P6/mmc$ (3)	24	2376	19300	$224328\frac{1}{8}$
$P6/mcm$, $P6/mmc$ (4)	24	2376	16300	$145928\frac{1}{8}$
Point group: 23				
$P23$, $P2_13$	12	276	760	$2695\frac{1}{4}$
$I23$, $I2_13$	24	2208	12160	86248
$F23$	48	17664	194560	2759936
Point group: $m\bar{3}$				
$Pm\bar{3}$, $Pn\bar{3}$, $Pa3$	24	1800	9400	$67703\frac{1}{8}$
$Im\bar{3}$, $Ia\bar{3}$	48	14400	150400	2166500
$Fm\bar{3}$	96	115200	2406400	69328000
$Fd\bar{3}$ (1)	96	115200	2406400	69328000
$Fd\bar{3}$ (2)	96	96768	1484800	28183680
Point group: 432				
$P432$, $P4_232$	24	1272	4648	$25216\frac{7}{8}$
$P4_132$† (15)	24	1272	4648	$25216\frac{7}{8}$
$P4_132$† (16)	24	1176	3568	$13916\frac{7}{8}$
$P4_132$† (17)	24	1080	2776	$8664\frac{7}{8}$
$P4_132$† (18)	24	984	2272	$6580\frac{7}{8}$
$I432$	48	10176	74368	806940
$I4_132$ (15)	48	10176	74368	806940
$I4_132$ (17)	48	8640	44416	277276
$F432$	96	81408	1189888	25822080
$F4_132$ (15)	96	81408	1189888	25822080
$F4_132$ (18)	96	62976	581632	6738816
Point group: $\bar{4}3m$				
$P\bar{4}3m$	24	1272	5128	$32896\frac{7}{8}$
$P\bar{4}3n$ (1)	24	1272	5128	$32896\frac{7}{8}$
$P\bar{4}3n$ (2)	24	1272	4168	$17536\frac{7}{8}$
$I\bar{4}3m$	48	10176	82048	1052700
$I\bar{4}3d$ (15); (20)	48	10176	82048	1052700
$I\bar{4}3d$ (15); (21)	48	10176	66688	561180
$I\bar{4}3d$ (17)	48	8640	44416	277276
$F\bar{4}3m$	96	81408	1312768	33686400
$F\bar{4}3c$ (15)	96	81408	1312768	33686400
$F\bar{4}3c$ (18)	96	81408	1067008	17957760
Point group: $m\bar{3}m$				
$Pm\bar{3}m$, $Pn\bar{3}m$	48	8784	72160	$972717\frac{13}{16}$
$Pn\bar{3}n$, $Pm\bar{3}n$ (1)	48	8784	72160	$972717\frac{13}{16}$
$Pn\bar{3}n$, $Pm\bar{3}n$ (2)	48	8784	56800	$488877\frac{13}{16}$
$Im\bar{3}m$	96	70272	1154560	31126970
$Ia\bar{3}d$ (15); (20)	96	70272	1154560	31126970
$Ia\bar{3}d$ (15); (21)	96	51840	432640	4497850
$Ia\bar{3}d$ (17)	96	70272	908800	15644090
$Fm\bar{3}m$	192	562176	18472960	996063040
$Fm\bar{3}c$ (1)	192	562176	18472960	996063040
$Fm\bar{3}c$ (2)	192	562176	14540800	500610880
$Fd\bar{3}m$ (1)	192	562176	18472960	996063040
$Fd\bar{3}m$ (2)	192	414720	7782400	205432640
$Fd\bar{3}c$ (1)	192	562176	18472960	996063040
$Fd\bar{3}c$ (2)	192	414720	6799360	136619840

Note. *hkl* subsets: (1) $h+k+l=2n$; (2) $h+k+l=2n+1$; (3) $l=2n$; (4) $l=2n+1$; (5) $2h+l=2n$; (6) $2h+l=2n+1$; (7) $2k+l=2n$; (8) $2k+l=2n+1$; (9) $l=6n$; (10) $l=6n+1, 6n+5$; (11) $l=6n+2, 6n+4$; (12) $l=6n+3$; (13) $l=3n$; (14) $l=3n+1$, $3n+2$; (15) *hkl* all even; (16) only one index odd; (17) only one index even; (18) *hkl* all odd; (19) two indices odd; (20) $h+k+l=4n$; (21) $h+k+l=4n+2$. † And the enantiomorphous space group.

Table 2.1.7.2. *Closed expressions for γ_{2k} [equation (2.1.7.11)] for space groups of low symmetry*

The normalized moments γ_{2k} are expressed in terms of M_k, where

$$M_k = \frac{(2k)!}{2^k(k!)^2} = \frac{(2k-1)!!}{k!},$$

and l', which takes on the values 1, 2 or 4 according as the Bravais lattice is of type P, one of the types A, B, C or I, or type F, respectively. The expressions for γ_{2k} are identical for all the space groups based on a given point group, except *Fdd*2 and *Fddd*. The expressions are valid for general reflections and under the restrictions given in the text.

Point group(s)	Expression for γ_{2k}
1	1
$\bar{1}$, 2, m	$l'^{k-1}M_k$
$2/m$, $mm2$	$l'^{k-1}M_k^2$
mmm	$l'^{k-1}M_k^3$
222	$\dfrac{l'^{k-1}}{2^k(k!)^2}\displaystyle\sum_{p=0}^{k}(M_pM_{k-p})^3[p!(k-p)!]^2$

$$f_k = n(x)He_k(2^{1/2}x), \qquad (2.1.7.16)$$

where $n(x)$ is the Gaussian distribution (Myller-Lebedeff, 1907). Unfortunately, no reasonably simple relationship between the coefficients d_k and readily evaluated properties of $p_c(|E|)$ has been found, and the Myller-Lebedeff expansion has not, as yet, been applied in crystallography. Although Stuart & Ord (1994, p. 112) dismiss it in a three-line footnote, it does have important applications in astronomy (van der Marel & Franx, 1993; Gerhard, 1993).

2.1.8. Non-ideal distributions: the Fourier method

The starting point of the method described in the previous section is the central-limit theorem approximation, and the method consists of finding correction factors which result in better approximations to the actual p.d.f. Conceptually, this is equivalent to improving the approximation of the characteristic function [*cf.* equation (2.1.4.10)] over that which led to the central-limit theorem result.

The method to be described in this section does not depend on any initial approximation and will be shown to utilize the dependence of the exact value of the characteristic function on the space-group symmetry, atomic composition and other factors. This approach has its origin in a simple but ingenious observation by Barakat (1974), who noted that if a random variable has lower and upper bounds then the corresponding p.d.f. can be nonzero only within these bounds and can therefore be expanded in an ordinary Fourier series and set to zero (identically) outside the bounded interval. Barakat's (1974) work dealt with intensity statistics of laser speckle, where sinusoidal waves are involved, as in the present problem. This method was applied by Weiss & Kiefer (1983) to testing the accuracy of a steepest-descents approximation to the exact solution of the problem of random walk, and its first application to crystallographic intensity statistics soon followed (Shmueli *et al.*, 1984). Crystallographic (*e.g.* Shmueli & Weiss, 1987; Rabinovich *et al.*, 1991*a*,*b*) and noncrystallographic (Shmueli *et al.*, 1985; Shmueli & Weiss, 1985*a*; Shmueli, Weiss & Wilson, 1989; Shmueli *et al.*, 1990) symmetry was found to be tractable by this approach, as well as joint conditional p.d.f.'s of several structure factors (Shmueli & Weiss, 1985*b*, 1986; Shmueli, Rabinovich & Weiss, 1989). The Fourier method is illustrated below by deriving the exact counterparts of equations (2.1.7.5) and (2.1.7.6) and specifying them for some simple symmetries. We shall then indicate a method of treating higher symmetries and present results which will suffice

for evaluation of Fourier p.d.f.'s of $|E|$ for a wide range of space groups.

2.1.8.1. General representations of p.d.f.'s of $|E|$ by Fourier series

We assume, as before, that (i) the atomic phase factors $\vartheta_j = 2\pi \mathbf{h}^T \mathbf{r}_j$ [*cf.* equation (2.1.1.2)] are uniformly distributed on (0–2π) and (ii) the atomic contributions to the structure factor are independent. For a centrosymmetric space group, with the origin chosen at a centre of symmetry, the random variable is the (real) normalized structure factor E and its bounds are $-E_{\rm M}$ and $E_{\rm M}$, where

$$E_{\rm M} = \sum_{j=1}^{N} n_j, \text{ with } n_j = \frac{f_j}{\left(\sum_{k=1}^{N} f_k^2\right)^{1/2}}. \tag{2.1.8.1}$$

Here, $E_{\rm M}$ is the maximum possible value of E and f_j is the conventional scattering factor of the jth atom, including its temperature factor. The p.d.f., $p(E)$, can be nonzero in the range $(-E_{\rm M}, E_{\rm M})$ only and can thus be expanded in the Fourier series

$$p(E) = (\alpha/2) \sum_{k=-\infty}^{\infty} C_k \exp(-\pi i k \alpha E), \tag{2.1.8.2}$$

where $\alpha = 1/E_{\rm M}$. Only the real part of $p(E)$ is relevant. The Fourier coefficients can be obtained in the conventional manner by integrating over the range $(-E_{\rm M}, E_{\rm M})$,

$$C_k = \int_{-E_{\rm M}}^{E_{\rm M}} p(E) \exp(\pi i k \alpha E)\, {\rm d}E. \tag{2.1.8.3}$$

Since, however, $p(E) = 0$ for $E < -E_{\rm M}$ and $E > E_{\rm M}$, it is possible and convenient to replace the limits of integration in equation (2.1.8.3) by infinity. Thus

$$C_k = \int_{-\infty}^{\infty} p(E) \exp(\pi i k \alpha E)\, {\rm d}E = \langle \exp(\pi i k \alpha E) \rangle. \tag{2.1.8.4}$$

Equation (2.1.8.4) shows that C_k is a Fourier transform of the p.d.f. $p(E)$ and, as such, it is the value of the corresponding characteristic function at the point $t_k = \pi \alpha k$ [*i.e.*, $C_k = C(\pi \alpha k$, where the characteristic function $C(t)$ is defined by equation (2.1.4.1)]. It is also seen that C_k is the expected value of the exponential $\exp(\pi i k \alpha E)$. It follows that the feasibility of the present approach depends on one's ability to evaluate the characteristic function in closed form *without* the knowledge of the p.d.f.; this is analogous to the problem of evaluating absolute moments of the structure factor for the correction-factor approach, discussed in Section 2.1.7. Fortunately, in crystallographic applications these calculations are feasible, provided individual isotropic motion is assumed. The formal expression for the p.d.f. of $|E|$, for any centrosymmetric space group, is therefore

$$p(|E|) = \alpha \left[1 + 2 \sum_{k=1}^{\infty} C_k \cos(\pi k \alpha |E|) \right], \tag{2.1.8.5}$$

where use is made of the assumption that $p(E) = p(-E)$, and the Fourier coefficients are evaluated from equation (2.1.8.4).

The p.d.f. of $|E|$ for a noncentrosymmetric space group is obtained by first deriving the joint p.d.f. of the real and imaginary parts of E and then integrating out its phase. The general expression for E is

$$E = A + iB = |E| \cos\varphi + i|E| \sin\varphi, \tag{2.1.8.6}$$

where φ is the phase of E. The required joint p.d.f. is

$$p(A, B) = (\alpha^2/4) \sum_m \sum_n C_{mn} \exp[-\pi i \alpha (mA + nB)], \tag{2.1.8.7}$$

and introducing polar coordinates $m = r \sin\Delta$ and $n = r \cos\Delta$, where $r = \sqrt{m^2 + n^2}$ and $\Delta = \tan^{-1}(m/n)$, we have

$$p(|E|, \varphi) = (\alpha^2/4)|E| \sum_m \sum_n C_{mn} \exp[-\pi i \alpha |E| \times \sqrt{m^2 + n^2} \sin(\varphi + \Delta)]. \tag{2.1.8.8}$$

Integrating out the phase φ, we obtain

$$p(|E|) = (\pi \alpha^2 |E|/2) \sum_m \sum_n C_{mn} J_0(\pi \alpha |E| \sqrt{m^2 + n^2}), \tag{2.1.8.9}$$

where $J_0(x)$ is the Bessel function of the first kind (*e.g.* Abramowitz & Stegun, 1972). This is a general form of the p.d.f. of $|E|$ for a noncentrosymmetric space group. The Fourier coefficients are obtained, similarly to the above, as

$$C_{mn} = \langle \exp[\pi i \alpha (mA + nB)] \rangle \tag{2.1.8.10}$$

and the average in equation (2.1.8.10), just as that in equation (2.1.8.4), is evaluated in terms of integrals over the appropriate trigonometric structure factors. In terms of the characteristic function for a joint p.d.f. of A and B, the Fourier coefficient in equation (2.1.8.10) is given by $C_{mn} = C(\pi \alpha m, \pi \alpha n)$.

We shall denote the characteristic function by $C(t_1)$ if it corresponds to a Fourier coefficient of a Fourier series for a centrosymmetric space group and by $C(t_1, t_2)$ or by $C(t, \Delta)$, where $t = (t_1^2 + t_2^2)^{1/2}$ and $\Delta = \tan^{-1}(t_1/t_2)$, if it corresponds to a Fourier series for a noncentrosymmetric space group.

2.1.8.2. Fourier–Bessel series

Equations (2.1.8.5) and (2.1.8.9) are the exact counterparts of equations (2.1.7.5) and (2.1.7.6), respectively. The computational effort required to evaluate equation (2.1.8.9) is somewhat greater than that for (2.1.8.5), because a double Fourier series has to be summed. The p.d.f. for *any* noncentrosymmetric space group can be expressed by a double Fourier series, but this can be simplified if the characteristic function depends on $t = (t_1^2 + t_2^2)^{1/2}$ alone, rather than on t_1 and t_2 separately. In such cases the p.d.f. of $|E|$ for a noncentrosymmetric space group can be expanded in a *single* Fourier–Bessel series (Barakat, 1974; Weiss & Kiefer, 1983; Shmueli *et al.*, 1984). The general form of this expansion is

$$p(|E|) = 2\alpha^2 |E| \sum_{u=1}^{\infty} D_u J_0(\alpha \lambda_u |E|), \tag{2.1.8.11}$$

where

$$D_u = \frac{C(\alpha \lambda_u)}{J_1^2(\lambda_u)} \tag{2.1.8.12}$$

and

$$C(\alpha \lambda_u) = \prod_{j=1}^{N/g} C_{ju}, \tag{2.1.8.13}$$

where $J_1(x)$ is the Bessel function of the first kind, and λ_u is the uth root of the equation $J_0(x) = 0$; the atomic contribution C_{ju} to equation (2.1.8.13) is computed as

$$C_{ju} = C(\alpha n_j \lambda_u). \tag{2.1.8.14}$$

The roots λ_u are tabulated in the literature (*e.g.* Abramowitz & Stegun, 1972), but can be most conveniently computed as follows. The first five roots are given by

$$\lambda_1 = 2.4048255577$$
$$\lambda_2 = 5.5200781103$$
$$\lambda_3 = 8.6537279129$$
$$\lambda_4 = 11.7915344390$$
$$\lambda_5 = 14.9309177085$$

and the higher ones can be obtained from McMahon's approximation (*cf.* Abramowitz & Stegun, 1972)

$$\lambda_u = \beta + \frac{1}{8\beta} - \frac{124}{3(8\beta)^3} + \frac{120928}{15(8\lambda)^5} - \frac{401743168}{105(8\lambda)^7} + \ldots, \quad (2.1.8.15)$$

where $\beta = (u - \frac{1}{4})\pi$. For $u > 5$ the values given by equation (2.1.8.15) have a relative error less than 10^{-11} so that no refinement of roots of higher orders is needed (Shmueli *et al.*, 1984). Numerical computations of single Fourier–Bessel series are of course faster than those of the double Fourier series, but both representations converge fairly rapidly.

2.1.8.3. Simple examples

Consider the Fourier coefficient of the p.d.f. of $|E|$ for the centrosymmetric space group $P\bar{1}$. The normalized structure factor is given by

$$E = 2\sum_{j=1}^{N/2} n_j \cos\vartheta_j, \quad \text{with} \quad \vartheta_j = 2\pi \mathbf{h}^T \cdot \mathbf{r}_j, \quad (2.1.8.16)$$

and the Fourier coefficient is

$$\begin{aligned} C_k &= \langle \exp(\pi i k \alpha E)\rangle & (2.1.8.17)\\ &= \left\langle \exp\left[2\pi i k\alpha \sum_{j=1}^{N/2} n_j \cos\vartheta_j\right]\right\rangle & (2.1.8.18)\\ &= \left\langle \prod_{j=1}^{N/2} \exp(2\pi i k\alpha n_j \cos\vartheta_j)\right\rangle & (2.1.8.19)\\ &= \prod_{j=1}^{N/2} \langle \exp(2\pi i k\alpha n_j \cos\vartheta_j)\rangle & (2.1.8.20)\\ &= \prod_{j=1}^{N/2} \left\{(1/2\pi)\int_{-\pi}^{\pi} \exp(2\pi i k\alpha n_j \cos\vartheta)\,\mathrm{d}\vartheta\right\} & (2.1.8.21)\\ &= \prod_{j=1}^{N/2} J_0(2\pi k\alpha n_j). & (2.1.8.22) \end{aligned}$$

Equation (2.1.8.20) is obtained from equation (2.1.8.19) if we make use of the assumption of independence, the assumption of uniformity allows us to rewrite equation (2.1.8.20) as (2.1.8.21), and the expression in the braces in the latter equation is just a definition of the Bessel function $J_0(2\pi k\alpha n_j)$ (*e.g.* Abramowitz & Stegun, 1972).

Let us now consider the Fourier coefficient of the p.d.f. of $|E|$ for the noncentrosymmetric space group $P1$. We have

$$A = \sum_{j=1}^{N} n_j \cos\vartheta_j \quad \text{and} \quad B = \sum_{j=1}^{N} n_j \sin\vartheta_j. \quad (2.1.8.23)$$

These expressions for A and B are substituted in equation (2.1.8.10), resulting in

$$\begin{aligned} C_{mn} &= \left\langle \prod_{j=1}^{N} \exp[\pi i\alpha n_j (m\cos\vartheta_j + n\sin\vartheta_j)]\right\rangle & (2.1.8.24)\\ &= \left\langle \prod_{j=1}^{N} \exp[\pi i\alpha n_j \sqrt{m^2+n^2}\sin(\vartheta_j + \Delta)]\right\rangle & (2.1.8.25)\\ &= \prod_{j=1}^{N} J_0(\pi\alpha n_j \sqrt{m^2+n^2}). & (2.1.8.26) \end{aligned}$$

Equation (2.1.8.24) leads to (2.1.8.25) by introducing polar coordinates analogous to those leading to equation (2.1.8.8), and equation (2.1.8.26) is then obtained by making use of the assumptions of independence and uniformity in an analogous manner to that detailed in equations (2.1.8.12)–(2.1.8.22) above.

The right-hand side of equation (2.1.8.26) is to be used as a Fourier coefficient of the double Fourier series given by (2.1.8.9). Since, however, this coefficient depends on $(m^2+n^2)^{1/2}$ alone rather than on m and n separately, the p.d.f. of $|E|$ for $P1$ can also be represented by a Fourier–Bessel series [*cf.* equation (2.1.8.11)] with coefficient

$$D_u = \frac{1}{J_1^2(\lambda_u)} \prod_{j=1}^{N} J_0(\alpha n_j \lambda_u), \quad (2.1.8.27)$$

where λ_u is the uth root of the equation $J_0(x) = 0$.

2.1.8.4. A more complicated example

We now illustrate the methodology of deriving characteristic functions for space groups of higher symmetries, following the method of Rabinovich *et al.* (1991*a*,*b*). The derivation is performed for the space group $P\bar{6}$ (No. 174). According to Table A1.4.3.6, the real and imaginary parts of the normalized structure factor are given by

$$\begin{aligned} A &= 2\sum_{j=1}^{N/6} n_j [C(hki)c(lz)]_j \\ &= 2\sum_{j=1}^{N/6} n_j \cos\tau_j \sum_{k=1}^{3} \cos\alpha_{jk} & (2.1.8.28) \end{aligned}$$

and

$$\begin{aligned} B &= 2\sum_{j=1}^{N/6} n_j [S(hki)c(lz)]_j \\ &= 2\sum_{j=1}^{N/6} n_j \cos\tau_j \sum_{k=1}^{3} \sin\alpha_{jk}, & (2.1.8.29) \end{aligned}$$

where

$$\begin{aligned} \alpha_{j1} &= 2\pi(hx_j + ky_j),\\ \alpha_{j2} &= 2\pi(kx_j + iy_j),\\ \alpha_{j3} &= 2\pi(ix_j + hy_j),\\ \tau_j &= 2\pi l z_j. \end{aligned}$$

Note that $\alpha_{j1} + \alpha_{j2} + \alpha_{j3} = 0$, *i.e.*, one of these contributions depends on the other two; this is a recurring problem in calculations pertaining to trigonal and hexagonal systems. For brevity, we write directly the general form of the characteristic function from which the functional form of the Fourier coefficient can be readily obtained. The characteristic function is given by

$$\begin{aligned} C(t_1, t_2) &= \langle \exp[i(t_1 A + t_2 B)]\rangle & (2.1.8.30)\\ &= \prod_{j=1}^{N/6} \left\langle \exp\left[2in_j \cos\tau_j \sum_{k=1}^{3} (t_1\cos\alpha_{jk} + t_2 \sin\alpha_{jk})\right]\right\rangle & (2.1.8.31)\\ &= \prod_{j=1}^{N/6} \left\langle \exp\left[2in_j t\cos\tau_j \sum_{k=1}^{3} (\sin\Delta\cos\alpha_{jk} \right.\right.\\ &\quad \left.\left. + \cos\Delta\sin\alpha_{jk})\right]\right\rangle & (2.1.8.32)\\ &= \prod_{j=1}^{N/6} \left\langle \exp\left[2in_j t\cos\tau_j \sum_{k=1}^{3} \sin(\alpha_{jk} + \Delta)\right]\right\rangle, & (2.1.8.33) \end{aligned}$$

where $\Delta = \tan^{-1}(t_1/t_2)$, $t = (t_1^2 + t_2^2)^{1/2}$ and the assumption of independence was used. If we further employ the assumption of

Table 2.1.8.1. *Atomic contributions to characteristic functions for $p(|E|)$*

The table lists symbolic expressions for the atomic contributions to exact characteristic functions (abbreviated as c.f.) for $p(|E|)$, to be computed as single Fourier series (centric), double Fourier series (acentric) and single Fourier–Bessel series (acentric), as defined in Sections 2.1.8.1 and 2.1.8.2. The symbolic expressions are defined in Section 2.1.8.5. The table is arranged by point groups, space groups and parities of the reflection indices analogously to the table of moments, Table 2.1.7.1, and covers all the space groups and statistically different parities of *hkl* up to and including space group $Fd\bar{3}$. The expressions are valid for atoms in general positions, for general reflections and presume the absence of noncrystallographic symmetry and of dispersive scatterers. The p.d.f.'s of $|E|$ for any centrosymmetric space group are computed from a single Fourier series [*cf.* (2.1.8.5)]. For a noncentrosymmetric space group the p.d.f. of $|E|$ *must* be computed from a double Fourier series [*cf.* (2.1.8.9)] – if the characteristic function in the table depends on Δ – and *may* be computed from a single Fourier–Bessel series [*cf.* (2.1.8.11)] if it does not depend on Δ. However, the p.d.f. of $|E|$ for any noncentrosymmetric space group may also be computed from a double Fourier series (*cf.* Section 2.1.8.1).

Space group(s)	g	Atomic c.f.	Remarks
Point group: 1			
$P1$	1	$J_0(tn_j)$	
Point group: $\bar{1}$			
$P\bar{1}$	2	$J_0(2t_1n_j)$	
Point groups: 2, *m*			
All P	2	$J_0^2(tn_j)$	
All C	4	$J_0^2(2tn_j)$	
Point group: 2/*m*			
All P	4	$J_0^2(2t_1n_j)$	
All C	8	$J_0^2(4t_1n_j)$	
Point group: 222			
All P	4	$L_j(t, \Delta)^{(a)}$	
All C and I	8	$L_j(2t, \Delta)$	
$F222$	16	$L_j(4t, \Delta)$	
Point group: *mm*2			
All P	4	$L_j(t, 0)$	
All C and I	8	$L_j(2t, 0)$	
$Fmm2$	16	$L_j(4t, 0)$	
$Fdd2$	16	$L_j(4t, 0)$	$h+k+l=2n$
	16	$L_j(4t, \pi/4)$	$h+k+l=2n+1$
Point group: *mmm*			
All P	8	$L_j(2t_1, 0)$	
All C and I	16	$L_j(4t_1, 0)$	
$Fmmm$	32	$L_j(8t_1, 0)$	
$Fddd$	32	$L_j(8t_1, 0)$	$h+k+l=2n$
	32	$L_j(8t_1, \pi/4)$	$h+k+l=2n+1$
Point group: 4			
$P4$, $P4_2$	4	$L_j(t, 0)$	
$P4_1$†	4	$L_j(t, 0)$	$l=2n$
	4	$L_j(t, \pi/4)$	$l=2n+1$
$I4$	8	$L_j(2t, 0)$	
$I4_1$	8	$L_j(2t, 0)$	$2h+l=2n$
	8	$L_j(2t, \pi/4)$	$2h+l=2n+1$
Point group: $\bar{4}$			
$P\bar{4}$	4	$L_j(t, \Delta)$	
$I\bar{4}$	8	$L_j(2t, \Delta)$	
Point group: 4/*m*			
All P	8	$L_j(2t_1, 0)$	
$I4/m$	16	$L_j(4t_1, 0)$	
$I4_1/a$	16	$L_j(4t_1, 0)$	$l=2n$
	16	$L_j(4t_1, \pi/4)$	$l=2n+1$
Point group: 422			
$P422$, $P42_12$, $P4_222$, $P4_22_12$	8	$Q_j^{(1)}(t, \Delta)^{(b)}$	
$P4_122$,† $P4_12_12$†	8	$Q_j^{(1)}(t, \Delta)$	$l=2n$
	8	$Q_j^{(2)}(t, \Delta)^{(c)}$	$l=2n+1$
$I422$	16	$Q_j^{(1)}(2t, \Delta)$	
$I4_122$	16	$Q_j^{(1)}(2t, \Delta)$	$2k+l=2n$
	16	$Q_j^{(2)}(2t, \Delta)$	$2k+l=2n+1$
Point group: 4*mm*			
All P	8	$Q_j^{(1)}(t, 0)$	
$I4mm$, $I4cm$	16	$Q_j^{(1)}(2t, 0)$	
$I4_1md$, $I4_1cd$	16	$Q_j^{(1)}(2t, 0)$	$2k+l=2n$
	16	$Q_j^{(1)}(2t, \pi/4)$	$2k+l=2n+1$
Point groups: $\bar{4}2m$, $\bar{4}m2$			
All P	8	$Q_j^{(1)}(t, \Delta)$	
$I\bar{4}2m$, $I\bar{4}m2$, $I\bar{4}c2$	16	$Q_j^{(1)}(2t, \Delta)$	
$I\bar{4}2d$	16	$Q_j^{(1)}(2t, \Delta)$	$2h+l=2n$
	16	$Q_j^{(2)}(2t, \Delta)$	$2h+l=2n+1$
Point group: 4/*mmm*			
All P	16	$Q_j^{(1)}(2t_1, 0)$	
$I4/mmm$, $I4/mcm$	32	$Q_j^{(1)}(4t_1, 0)$	
$I4_1/amd$, $I4_1/acd$	32	$Q_j^{(1)}(4t_1, 0)$	$l=2n$
	32	$Q_j^{(1)}(4t_1, \pi/4)$	$l=2n+1$
Point group: 3			
All P and R	3	$J_0^3(m_j)$	
Point group: $\bar{3}$			
All P and R	6	$J_0^3(2t_1n_j)$	
Point group: 32			
All P and R	6	$T_j(t, \Delta)^{(d)}$	
Point group: 3*m*			
$P3m1$, $P31m$, $R3m$	6	$T_j(t, \pi/2)$	
$P3c1$, $P31c$, $R3c$	6	$T_j(t, \pi/2)$	$l=2n\,(P)$, $h+k+l=2n\,(R)$
	6	$T_j(t, 0)$	$l=2n+1\,(P)$, $h+k+l=2n+1\,(R)$
Point group: $\bar{3}m$			
$P\bar{3}m1$, $P\bar{3}1m$, $R\bar{3}m$	12	$T_j(2t_1, \pi/2)$	
$P\bar{3}c1$, $P\bar{3}1c$, $R\bar{3}c$	12	$T_j(2t_1, \pi/2)$	$l=2n\,(P)$, $h+k+l=2n\,(R)$
	12	$T_j(2t_1, 0)$	$l=2n+1\,(P)$, $h+k+l=2n+1\,(R)$
Point group: 6			
$P6$	6	$H_j^{(1)}(t, \pi/2)^{(e)}$	
$P6_1$†	6	$H_j^{(1)}(t, \pi/2)$	$l=6n$
	6	$H_j^{(2)}(t, 0)^{(f)}$	$l=6n+1, 6n+5$
	6	$H_j^{(2)}(t, \pi/2)$	$l=6n+2, 6n+4$
	6	$H_j^{(1)}(t, 0)$	$l=6n+3$
$P6_2$†	6	$H_j^{(1)}(t, \pi/2)$	$l=3n$
	6	$H_j^{(2)}(t, \pi/2)$	$l=3n\pm1$
$P6_3$	6	$H_j^{(1)}(t, \pi/2)$	$l=2n$
	6	$H_j^{(1)}(t, 0)$	$l=2n+1$
Point group: $\bar{6}$			
$P\bar{6}$	6	$H_j^{(1)}(t, \Delta)$	
Point group: 6/*m*			
$P6/m$	12	$H_j^{(1)}(2t_1, \pi/2)$	
$P6_3/m$	12	$H_j^{(1)}(2t_1, \pi/2)$	$l=2n$
	12	$H_j^{(1)}(2t_1, 0)$	$l=2n+1$
Point group: 622			
$P622$	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)^{(g)}$	
$P6_122$†	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)$	$l=6n$
	12	$\tilde{H}_j^{(2)}(t, 0, 0, \Delta)^{(h)}$	$l=6n+1, 6n+5$
	12	$\tilde{H}_j^{(2)}(t, \pi/2, \pi/2, \Delta)$	$l=6n+2, 6n+4$
	12	$\tilde{H}_j^{(1)}(t, 0, 0, \Delta)$	$l=6n+3$
$P6_222$†	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)$	$l=3n$
	12	$\tilde{H}_j^{(2)}(t, \pi/2, \pi/2, \Delta)$	$l=3n\pm1$

Table 2.1.8.1 (*cont.*)

Space group(s)	g	Atomic c.f.	Remarks
$P6_322$	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, 0, 0, \Delta)$	$l = 2n+1$
Point group: 6*mm*			
$P6mm$	12	$\tilde{H}_j^{(1)}(t, \pi/2, \pi/2, 0)$	
$P6cc$	12	$\tilde{H}_j^{(1)}(t, \pi/2, \pi/2, 0)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, 0)$	$l = 2n+1$
$P6_3cm$, $P6_3mc$	12	$\tilde{H}_j^{(1)}(t, \pi/2, \pi/2, 0)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, 0, 0, 0)$	$l = 2n+1$
Point groups: $\bar{6}2m$, $\bar{6}m2$			
$P\bar{6}2m$, $P\bar{6}m2$	12	$\tilde{H}_j^{(1)}(t, \Delta, \Delta, 0)$	
$P\bar{6}2c$, $P\bar{6}c2$	12	$\tilde{H}_j^{(1)}(t, \Delta, \Delta, 0)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, \Delta+\pi/2, -\Delta-\pi/2, 0)$	$l = 2n+1$
Point group: 6/*mmm*			
$P6/mmm$	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, \pi/2, 0)$	
$P6/mcc$	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, \pi/2, 0)$	$l = 2n$
	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, -\pi/2, 0)$	$l = 2n+1$
$P6_3/mcm$, $P6_3/mmc$	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, \pi/2, 0)$	$l = 2n$
	24	$\tilde{H}_j^{(1)}(2t_1, 0, 0, 0)$	$l = 2n+1$
Point group: 23			
$P23$, $P2_13$	12	$L_j^3(t, \Delta)$	
$I23$, $I2_13$	24	$L_j^3(2t, \Delta)$	
$F23$	48	$L_j^3(4t, \Delta)$	
Point group: $m\bar{3}$			
$Pm\bar{3}$, $Pn\bar{3}$, $Pa\bar{3}$	24	$L_j^3(2t_1, 0)$	
$Im\bar{3}$, $Ia\bar{3}$	48	$L_j^3(4t_1, 0)$	
$Fm\bar{3}$	96	$L_j^3(8t_1, 0)$	
$Fd\bar{3}$	96	$L_j^3(8t_1, 0)$	$h+k+l = 2n$
	96	$L_j^3(8t_1, \pi/4)$	$h+k+l = 2n+1$

† And the enantiomorphous space group.

uniformity, while remembering that the angular variables α_{jk} are not independent, the characteristic function can be written as

$$C(t_1, t_2) = \prod_{j=1}^{N/6}\left\{(1/2\pi)\int_{-\pi}^{\pi} \mathrm{d}\tau\left([1/(2\pi)^2]\right.\right. \times \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathrm{d}\alpha_1\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_3\,\delta_{2\pi}(\alpha_1+\alpha_2+\alpha_3) \left.\left. \times \exp\left[2in_jt\cos\tau\sum_{k=1}^{3}\sin(\alpha_k+\Delta)\right]\right)\right\}, \tag{2.1.8.34}$$

where

$$\delta_{2\pi}(\alpha) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\exp(-ik\alpha) \tag{2.1.8.35}$$

is the Fourier representation of the periodic delta function. Equation (2.1.8.34) then becomes

$$C(t_1, t_2) = \prod_{j=1}^{N/6}\left\{(1/2\pi)\int_{-\pi}^{\pi}\mathrm{d}\tau\sum_{k=-\infty}^{\infty}\left[(1/2\pi) \times \int_{-\pi}^{\pi}\exp\bigl(-ik\alpha+2in_jt\cos\tau\sin(\alpha+\Delta)\bigr)\,\mathrm{d}\alpha\right]^3\right\}. \tag{2.1.8.36}$$

If we change the variable α to $\alpha'-\Delta$, $\sin(\alpha+\Delta)$ becomes $\sin\alpha'$ and $-ik\alpha = -ik\alpha' + ik\Delta$. Hence

$$C(t_1, t_2) = \prod_{j=1}^{N/6}\left\{(1/2\pi)\int_{-\pi}^{\pi}\mathrm{d}\tau\sum_{k=-\infty}^{\infty}\exp(3ik\Delta)J_k^3(2n_jt\cos\tau)\right\}. \tag{2.1.8.37}$$

The imaginary part of the summation, involving Bessel functions of odd orders, vanishes upon integration and the latter is restricted to the positive quadrant in τ. Thus, upon replacing cosines by sines (this is permissible at this stage) the atomic contribution to the characteristic function becomes

$$C_j(t, \Delta) = (2/\pi)\int_0^{\pi/2}\left[J_0^3(2n_jt\sin\tau) + 2\sum_{k=1}^{\infty}\cos(6k\Delta)J_k^3(2n_jt\sin\tau)\right]\mathrm{d}\tau \tag{2.1.8.38}$$

and a double Fourier series must be used for the p.d.f.

2.1.8.5. Atomic characteristic functions

Expressions for the atomic contributions to the characteristic functions were obtained by Rabinovich *et al.* (1991*a*) for a wide range of space groups, by methods similar to those described above. These expressions are collected in Table 2.1.8.1 in terms of symbols which are defined below. The following abbreviations are used in the subsequent definitions of the symbols:

$$s_\pm = 2an_j\sin(\tau\pm\rho),$$
$$c_\pm = 2an_j\cos(\tau\pm\rho) \quad \text{and}$$
$$\sigma_\pm = 2an_j\sin(\tau\pm 2\pi/3+\rho),$$

and the symbols appearing in Table 2.1.8.1 are given below:

$$\begin{aligned}
{}^{(a)}L_j(a,\rho) &= \langle J_0(s_+)J_0(s_-)\rangle_\tau \\
&= \sum_{k=-\infty}^{\infty}\cos(4k\rho)J_k^4 \\
&= J_0^4(an_j) + 2\sum_{k=1}^{\infty}\cos(4k\rho)J_k^4(an_j), \\
{}^{(b)}Q_j^{(1)}(a,\rho) &= \langle J_0^2(s_+)J_0^2(s_-)\rangle_\tau, \\
{}^{(c)}Q_j^{(2)}(a,\rho) &= \langle J_0(s_+)J_0(s_-)J_0(c_+)J_0(c_-)\rangle_\tau, \\
{}^{(d)}T_j(a,\rho) &= \sum_{k=-\infty}^{\infty}\exp(6ik\rho)J_k^6(an_j) \\
&= J_0^6(an_j) + 2\sum_{k=1}^{\infty}\cos(6k\rho)J_k^6(an_j), \\
{}^{(e)}H_j^{(1)}(a,\mu) &= \left\langle\mathcal{R}\left[S_j^{(1)}(\tau;\ a,\mu,0)\right]\right\rangle_\tau, \\
{}^{(f)}H_j^{(2)}(a,\mu) &= \left\langle\mathcal{R}\left[S_j^{(2)}(\tau;\ a,\mu,0)\right]\right\rangle_\tau, \\
{}^{(g)}\tilde{H}_j^{(1)}(a,\mu_1,\mu_2,\rho) &= \left\langle\mathcal{R}\left[S_j^{(1)}(\tau;\ a,\mu_1,\rho) \times S_j^{(1)}(\tau;\ a,\mu_2,-\rho)\right]\right\rangle_\tau, \\
{}^{(h)}\tilde{H}_j^{(2)}(a,\mu_1,\mu_2,\rho) &= \left\langle\mathcal{R}\left[S_j^{(2)}(\tau;\ a,\mu_1,\rho) \times S_j^{(2)}(\tau;\ a,\mu_2,-\rho)\right]\right\rangle_\tau,
\end{aligned}$$

where

$$S_j^{(1)}(\tau;\ a, \mu, \rho) = \sum_{k=-\infty}^{\infty} e^{3ik\mu} J_k^3(s_+)$$

and

$$S_j^{(2)}(\tau;\ a, \mu, \rho) = \sum_{k=-\infty}^{\infty} e^{3ik\mu} J_k(s_+) J_k(\sigma_+) J_k(\sigma_-).$$

The averages appearing in the above summary are, in general, computed as

$$\langle f(\tau) \rangle = (2/\pi) \int_0^{\pi/2} f(\tau)\, \mathrm{d}\tau, \qquad (2.1.8.39)$$

except $H_j^{(2)}$ and $\tilde{H}_j^{(2)}$ which are computed as

$$\langle f(\tau) \rangle = (3/\pi) \int_0^{\pi/3} f(\tau)\, \mathrm{d}\tau, \qquad (2.1.8.40)$$

where $f(\tau)$ is any of the atomic characteristic functions indicated above. The superscripts preceding the symbols in the above summary are appended to the corresponding symbols in Table 2.1.8.1 on their first occurrence.

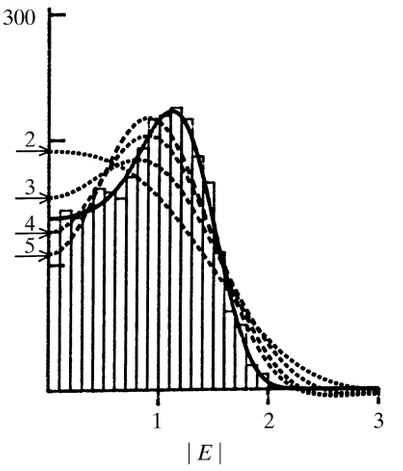


Fig. 2.1.8.1. The same recalculated histogram as in Fig. 2.1.7.1 along with the centric correction-factor p.d.f. [equation (2.1.7.5)], truncated after two, three, four and five terms (dashed lines), and with that accurately computed for the correct space-group Fourier p.d.f. [equations (2.1.8.5) and (2.1.8.22)] (solid line).

2.1.8.6. Other non-ideal Fourier p.d.f.'s

As pointed out above, the representation of the p.d.f.'s by Fourier series is also applicable to effects of noncrystallographic symmetry. Thus, Shmueli *et al.* (1985) obtained the following Fourier coefficient for the bicentric distribution in the space group $P\bar{1}$:

$$C_k = (2/\pi) \int_0^{\pi/2} \left[\prod_{j=1}^{N/4} J_0(4\pi k \alpha n_j \cos\vartheta) \right] \mathrm{d}\vartheta, \qquad (2.1.8.41)$$

to be used with equation (2.1.8.5). Furthermore, if we use the important property of the characteristic function as outlined in Section 2.1.4.1, it is easy to write down the Fourier coefficient for a $P\bar{1}$ asymmetric unit containing a centrosymmetric fragment centred at a noncrystallographic centre and a number of atoms not related by symmetry. This Fourier for the above partially bicentric arrangement is a product of expressions (2.1.8.17) and (2.1.8.41), with the appropriate number of atoms in each factor (Shmueli & Weiss, 1985*a*). While the purely bicentric p.d.f. obtained by using (2.1.8.41) with (2.1.8.5) is significantly different from the ideal bicentric p.d.f. given by equation (2.1.5.13) only when the atomic composition is sufficiently heterogeneous, the above partially bicentric p.d.f. appears to be a useful development even for an equal-atom structure.

The problem of the coexistence of several noncrystallographic centres of symmetry within the asymmetric unit of $P\bar{1}$, and its effect on the p.d.f. of $|E|$, was examined by Shmueli, Weiss & Wilson (1989) by the Fourier method. The latter study indicates that the strongest effect is produced by the presence of a single noncrystallographic centre.

Another kind of noncrystallographic symmetry is that arising from the presence of centrosymmetric fragments in a noncentrosymmetric structure – the subcentric arrangement already discussed in Section 2.1.5.4. A Fourier-series representation of a non-ideal p.d.f. corresponding to this case was developed by Shmueli, Rabinovich & Weiss (1989), and was also applied to the mathematically equivalent effects of dispersion and presence of heavy scatterers in centrosymmetric special positions in a noncentrosymmetric space group.

A variety of other non-ideal p.d.f.'s occur when heavy atoms are present in special positions (Shmueli & Weiss, 1988). Without going into the details of this development, it can be noted that if the atoms are distributed among k types of Wyckoff positions, the characteristic function corresponding to the p.d.f. of $|E|$ is a product of the k characteristic functions, each of which is related to one of these special positions; the same property of the characteristic function as that in Section 2.1.4.1 is here utilized.

2.1.8.7. Comparison of the correction-factor and Fourier approaches

The need for theoretical non-ideal distributions was exemplified by Fig. 2.1.7.1, referred to above, and the performance of the two approaches described above, for this particular example, is shown in Fig. 2.1.8.1. Briefly, the Fourier p.d.f. shows an excellent agreement with the histogram of recalculated $|E|$ values, while the agreement attained by the Hermite correction factor is much less satisfactory, even for the (longest available to us) five-term expansion. It must be pointed out that (i) the inadequacy of 'short' correction factors, in the example shown, is due to the large deviation from the ideal behaviour and (ii) the number of terms used there in the Fourier summation is twenty, whereafter the summation is terminated. Obviously, the computation of twenty (or more) Fourier coefficients is easier than that of five terms in the correction factor. The convergence of the Fourier series is very satisfactory. It appears that the (analytically) exact Fourier approach is the preferred one in cases of large or intermediate deviations, while the correction-factor approach may cope well with small ones. As far as the availability of symmetry-dependent centric and acentric p.d.f.'s is concerned, correction factors are available for all the space groups (see Table 2.1.7.1), while Fourier coefficients of p.d.f.'s are available for the first 206 space groups (see Table 2.1.8.1). It should be pointed out that p.d.f.'s based on the correction-factor method cope very well with cubic symmetries higher than $Fd\bar{3}$, even if the asymmetric unit of the space group is strongly heterogeneous (Rabinovich *et al.*, 1991*b*).

Both approaches described in this section are related to the characteristic function of the required p.d.f. The correction-factor p.d.f.'s (2.1.7.5) and (2.1.7.6) can be obtained by expanding the logarithm of the appropriate characteristic function in a series of cumulants [*e.g.* equation (2.1.4.13); see also Shmueli & Wilson (1982)], truncating the series and performing its term-by-term Fourier inversion. The Fourier p.d.f., on the other hand, is computed by forming a Fourier series whose coefficients are *exact* analytical forms of the characteristic function at points related to the summation indices [*e.g.* equations (2.1.8.5), (2.1.8.9) and (2.1.8.11), and Table 2.1.8.1] and truncating the series when the terms become small enough.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.

Barakat, R. (1974). *First-order statistics of combined random sinusoidal waves with application to laser speckle patterns. Opt. Acta*, **21**, 903–921.

Bernstein, S. (1922). *Sur la théorème limite du calcul des probabilités. Math. Ann.* **85**, 237–241.

Bernstein, S. (1927). *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann.* **97**, 1–59.

Cramér, H. (1951). *Mathematical Methods of Statistics.* Princeton University Press.

Faggiani, R., Lippert, B. & Lock, C. J. L. (1980). *Heavy transition metal complexes of biologically important molecules. 4. Crystal and molecular structure of pentahydroxonium chloro(uracilato-N(1))(ethylenediamine)platinum(II)chloride* $(H_5O_2)[PtCl(NH_2CH_2CH_2NH_2)(C_4H_5N_2O_2)]Cl$, *and chloro(thyminato-N(1))(ethylenediamine)platinum(II),* $PtCl(NH_2CH_2CH_2NH_2)(C_5H_5N_2O_2)$. *Inorg. Chem.* **19**, 295–300.

French, S. & Wilson, K. (1978). *On the treatment of negative intensity observations. Acta Cryst.* A**34**, 517–525.

Gerhard, O. E. (1993). *Line-of-sight velocity profiles in spherical galaxies: breaking the degeneracy between anisotropy and mass. Mon. Not. R. Astron. Soc.* **265**, 213–230.

Giacovazzo, C. (1977). *On different probabilistic approaches to quartet theory. Acta Cryst.* A**33**, 50–54.

Giacovazzo, C. (1980). *Direct Methods in Crystallography.* London: Academic Press.

Harker, D. (1953). *The meaning of the average of* $|F|^2$ *for large values of interplanar spacing. Acta Cryst.* **6**, 731–736.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* Am. Crystallogr. Assoc. Monograph No. 3. Dayton, Ohio: Polycrystal Book Service.

Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *The probability distribution of X-ray intensities. II. Experimental investigation and the X-ray detection of centers of symmetry. Acta Cryst.* **3**, 210–214.

International Tables for Crystallography (2005). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn. Heidelberg: Springer.

International Tables for Crystallography (2004). Vol. C, *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

Kendall, M. & Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. London: Griffin.

Lipson, H. & Woolfson, M. M. (1952). *An extension of the use of intensity statistics. Acta Cryst.* **5**, 680–682.

Lomer, T. R. & Wilson, A. J. C. (1975). *Scaling of intensities. Acta Cryst.* B**31**, 646–647.

Marel, R. P. van der & Franx, M. (1993). *A new method for the identification of non-Gaussian line profiles in elliptical galaxies. Astrophys. J.* **407**, 525–539.

Myller-Lebedeff, W. (1907). *Die Theorie der Integralgleichungen in Anwendung auf einige Reihenentwicklungen. Math. Ann.* **64**, 388–416.

Nigam, G. D. (1972). *On the compensation of X-ray intensity. Indian J. Pure Appl. Phys.* **10**, 655–656.

Nigam, G. D. & Wilson, A. J. C. (1980). *Compensation of excess intensity in space group P2. Acta Cryst.* A**36**, 832–833.

Rabinovich, D. & Shakked, Z. (1984). *A new approach to structure determination of large molecules by multi-dimensional search methods. Acta Cryst.* A**40**, 195–200.

Rabinovich, S., Shmueli, U., Stein, Z., Shashua, R. & Weiss, G. H. (1991*a*). *Exact random-walk models in crystallographic statistics. VI. P.d.f.'s of* $|E|$ *for all plane groups and most space groups. Acta Cryst.* A**47**, 328–335.

Rabinovich, S., Shmueli, U., Stein, Z., Shashua, R. & Weiss, G. H. (1991*b*). *Exact random-walk models in crystallographic statistics. VII. An all-space-group study of the effects of atomic heterogeneity on the p.d.f.'s of* $|E|$. *Acta Cryst.* A**47**, 336–340.

Rayleigh, Lord (1879). *Investigations in optics with special reference to the spectroscope. Philos. Mag.* **8**, 261–274.

Rogers, D. (1950). *The probability distribution of X-ray intensities. IV. New methods of determining crystal classes and space groups. Acta Cryst.* **3**, 455–464.

Rogers, D. & Wilson, A. J. C. (1953). *The probability distribution of X-ray intensities. V. A note on some hypersymmetric distributions. Acta Cryst.* **6**, 439–449.

Shmueli, U. (1979). *Symmetry- and composition-dependent cumulative distributions of the normalized structure amplitude for use in intensity statistics. Acta Cryst.* A**35**, 282–286.

Shmueli, U. (1982). *A study of generalized intensity statistics: extension of the theory and practical examples. Acta Cryst.* A**38**, 362–371.

Shmueli, U. & Kaldor, U. (1981). *Calculation of even moments of the trigonometric structure factor. Methods and results. Acta Cryst.* A**37**, 76–80.

Shmueli, U. & Kaldor, U. (1983). *Moments of the trigonometric structure factor. Acta Cryst.* A**39**, 615–621.

Shmueli, U., Rabinovich, S. & Weiss, G. H. (1989). *Exact conditional distribution of a three-phase invariant in the space group P1. I. Derivation and simplification of the Fourier series. Acta Cryst.* A**45**, 361–367.

Shmueli, U., Rabinovich, S. & Weiss, G. H. (1990). *Exact random-walk models in crystallographic statistics. V. Non-symmetrically bounded distributions of structure-factor magnitudes. Acta Cryst.* A**46**, 241–246.

Shmueli, U. & Weiss, G. H. (1985*a*). *Centric, bicentric and partially bicentric intensity statistics*. In *Structure and Statistics in Crystallography*, edited by A. J. C. Wilson, pp. 53–66. Guilderland: Adenine Press.

Shmueli, U. & Weiss, G. H. (1985*b*). *Exact joint probability distributions for centrosymmetric structure factors. Derivation and application to the* Σ_1 *relationship in the space group* $P\bar{1}$. *Acta Cryst.* A**41**, 401–408.

Shmueli, U. & Weiss, G. H. (1986). *Exact joint distribution of* $E_{\boldsymbol{h}}$, $E_{\boldsymbol{k}}$ *and* $E_{\boldsymbol{h}+\boldsymbol{k}}$ *and the probability for the positive sign of the triple product in the space group* $P\bar{1}$. *Acta Cryst.* A**42**, 240–246.

Shmueli, U. & Weiss, G. H. (1987). *Exact random-walk models in crystallographic statistics. III. Distributions of* $|E|$ *for space groups of low symmetry. Acta Cryst.* A**43**, 93–98.

Shmueli, U. & Weiss, G. H. (1988). *Exact random-walk models in crystallographic statistics. IV. P.d.f.'s of* $|E|$ *allowing for atoms in special positions. Acta Cryst.* A**44**, 413–417.

Shmueli, U., Weiss, G. H. & Kiefer, J. E. (1985). *Exact random-walk models in crystallographic statistics. II. The bicentric distribution in the space group* $P\bar{1}$. *Acta Cryst.* A**41**, 55–59.

Shmueli, U., Weiss, G. H., Kiefer, J. E. & Wilson, A. J. C. (1984). *Exact random-walk models in crystallographic statistics. I. Space groups* $P\bar{1}$ *and* $P1$. *Acta Cryst.* A**40**, 651–660.

Shmueli, U., Weiss, G. H. & Wilson, A. J. C. (1989). *Explicit Fourier representations of non-ideal hypercentric p.d.f.'s of* $|E|$. *Acta Cryst.* A**45**, 213–217.

Shmueli, U. & Wilson, A. J. C. (1981). *Effects of space-group symmetry and atomic heterogeneity on intensity statistics. Acta Cryst.* A**37**, 342–353.

Shmueli, U. & Wilson, A. J. C. (1982). *Intensity statistics: non-ideal distributions in theory and practice.* In *Crystallographic Statistics: Progress and Problems*, edited by S. Ramaseshan, M. F. Richardson & A. J. C. Wilson, pp. 83–97. Bangalore: Indian Academy of Sciences.

Shmueli, U. & Wilson, A. J. C. (1983). *Generalized intensity statistics: the subcentric distribution and effects of dispersion. Acta Cryst.* A**39**, 225–233.

Spiegel, M. R. (1974). *Theory and Problems of Fourier Analysis.* Schaum's Outline Series. New York: McGraw-Hill.

Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.

Stuart, A. & Ord, K. (1994). *Kendall's Advanced Theory of Statistics*, Vol. 1, *Distribution Theory*, 6th ed. London: Edward Arnold.

Weiss, G. H. & Kiefer, J. E. (1983). *The Pearson random walk with unequal step sizes. J. Phys. A*, **16**, 489–495.

Wilson, A. J. C. (1942). *Determination of absolute from relative intensity data. Nature (London)*, **150**, 151–152.

Wilson, A. J. C. (1949). *The probability distribution of X-ray intensities. Acta Cryst.* **2**, 318–320.

Wilson, A. J. C. (1950). *The probability distribution of X-ray intensities. III. Effects of symmetry elements on zones and rows. Acta Cryst.* **3**, 258–261.

Wilson, A. J. C. (1952). *Hypercentric and hyperparallel distributions of X-ray intensities. Research (London)*, **5**, 588–589.

Wilson, A. J. C. (1956). *The probability distribution of X-ray intensities. VII. Some sesquicentric distributions. Acta Cryst.* **9**, 143–144.

Wilson, A. J. C. (1964). *The probability distribution of X-ray intensities. VIII. A note on compensation for excess average intensity. Acta Cryst.* **17**, 1591–1592.

Wilson, A. J. C. (1975). *Effect of neglect of dispersion on apparent scale and temperature factors.* In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 325–332. Copenhagen: Munksgaard.

Wilson, A. J. C. (1976). *Statistical bias in least-squares refinement. Acta Cryst.* A**32**, 994–996.

Wilson, A. J. C. (1978*a*). *On the probability of measuring the intensity of a reflection as negative. Acta Cryst.* A**34**, 474–475.

Wilson, A. J. C. (1978*b*). *Variance of X-ray intensities: effect of dispersion and higher symmetries. Acta Cryst.* A**34**, 986–994.

Wilson, A. J. C. (1978*c*). *Statistical bias in scaling factors: Erratum. Acta Cryst.* B**34**, 1749.

Wilson, A. J. C. (1979). *Problems of resolution and bias in the experimental determination of the electron density and other densities in crystals. Acta Cryst.* A**35**, 122–130.

Wilson, A. J. C. (1980*a*). *Relationship between 'observed' and 'true' intensity: effects of various counting modes. Acta Cryst.* A**36**, 929–936.

Wilson, A. J. C. (1980*b*). *Effect of dispersion on the probability distribution of X-ray reflections. Acta Cryst.* A**36**, 945–946.

Wilson, A. J. C. (1981). *Can intensity statistics accommodate stereochemistry? Acta Cryst.* A**37**, 808–810.

Wilson, A. J. C. (1986*a*). *Distributions of sums and ratios of sums of intensities. Acta Cryst.* A**42**, 334–339.

Wilson, A. J. C. (1986*b*). *Fourier versus Hermite representations of probability distributions. Acta Cryst.* A**42**, 81–83.

Wilson, A. J. C. (1987*a*). *Treatment of enhanced zones and rows in normalizing intensities. Acta Cryst.* A**43**, 250–252.

Wilson, A. J. C. (1987*b*). *Functional form of the ideal hypersymmetric distributions of structure factors. Acta Cryst.* A**43**, 554–556.

Wilson, A. J. C. (1993). *Space groups rare for organic structures. III. Symmorphism and inherent symmetry. Acta Cryst.* A**49**, 795–806.

International Tables for Crystallography (2010). Vol. B, Chapter 2.2, pp. 215–243.

2.2. Direct methods

By C. Giacovazzo

2.2.1. List of symbols and abbreviations

f_j	atomic scattering factor of jth atom
Z_j	atomic number of jth atom
N	number of atoms in the unit cell
m	order of the point group

$$[\sigma_r]_p, [\sigma_r]_q, [\sigma_r]_N, \ldots = \sum_{j=1}^{p} Z_j^r, \sum_{j=1}^{q} Z_j^r, \sum_{j=1}^{N} Z_j^r, \ldots$$

$[\sigma_r]_N$ is always abbreviated to σ_r when N is the number of atoms in the cell

$$\textstyle\sum_p, \sum_q, \sum_N, \ldots = \sum_{j=1}^{p} f_j^2, \sum_{j=1}^{q} f_j^2, \sum_{j=1}^{N} f_j^2, \ldots$$

s.f.	structure factor
n.s.f.	normalized structure factor
cs.	centrosymmetric
ncs.	noncentrosymmetric
s.i.	structure invariant
s.s.	structure seminvariant
$\mathbf{C} = (\mathbf{R}, \mathbf{T})$	symmetry operator; $\mathbf{R}$ is the rotational part, $\mathbf{T}$ the translational part
$\varphi_{\mathbf{h}}$	phase of the structure factor $F_{\mathbf{h}} = \lvert F_{\mathbf{h}}\rvert \exp(i\varphi_{\mathbf{h}})$

2.2.2. Introduction

Direct methods are today the most widely used tool for solving small crystal structures. They work well both for equal-atom molecules and when a few heavy atoms exist in the structure. In recent years the theoretical background of direct methods has been improved to take into account a large variety of prior information (the form of the molecule, its orientation, a partial structure, the presence of pseudosymmetry or of a superstructure, the availability of isomorphous data or of data affected by anomalous-dispersion effects, ...). Owing to this progress and to the increasing availability of powerful computers, the phase problem for small molecules has been solved in practice: a number of effective, highly automated packages are today available to the scientific community.

The combination of direct methods with so-called direct-space methods have recently allowed the *ab initio* crystal structure solution of proteins. The present limit of complexity is about 2500 non-hydrogen atoms in the asymmetric unit, but diffraction data at atomic resolution (~1 Å) are required. Trials are under way to bring this limit to 1.5 Å and have shown some success.

The theoretical background and tables useful for origin specification are given in Section 2.2.3; in Section 2.2.4 the procedures for normalizing structure factors are summarized. Phase-determining formulae (inequalities, probabilistic formulae for triplet, quartet and quintet invariants, and for one- and two-phase s.s.'s, determinantal formulae) are given in Section 2.2.5. In Section 2.2.6 the connection between direct methods and related techniques in real space is discussed. Practical procedures for solving small-molecule crystal structures are described in Sections 2.2.7 and 2.2.8, and references to the most extensively used packages are given in Section 2.2.9. The integration of direct methods, isomorphous replacement and anomalous-dispersion techniques is briefly discussed in Section 2.2.10.

The reader interested in a more detailed description of the topic is referred to a recent textbook (Giacovazzo, 1998).

2.2.3. Origin specification

(*a*) Once the origin has been chosen, the symmetry operators $\mathbf{C}_s \equiv (\mathbf{R}_s, \mathbf{T}_s)$ and, through them, the algebraic form of the s.f. remain fixed.

A shift of the origin through a vector with coordinates $\mathbf{X}_0$ transforms $\varphi_{\mathbf{h}}$ into

$$\varphi'_{\mathbf{h}} = \varphi_{\mathbf{h}} - 2\pi\mathbf{h}\cdot\mathbf{X}_0 \qquad (2.2.3.1)$$

and the symmetry operators $\mathbf{C}_s$ into $\mathbf{C}'_s = (\mathbf{R}'_s, \mathbf{T}'_s)$, where

$$\mathbf{R}'_s = \mathbf{R}_s;\ \ \mathbf{T}'_s = \mathbf{T}_s + (\mathbf{R}_s - \mathbf{I})\mathbf{X}_0 \quad s = 1, 2, \ldots, m. \qquad (2.2.3.2)$$

(*b*) *Allowed* or *permissible origins* (Hauptman & Karle, 1953, 1959) for a given algebraic form of the s.f. are all those points in direct space which, when taken as origin, maintain the same symmetry operators $\mathbf{C}_s$. The allowed origins will therefore correspond to those points having the same symmetry environment in the sense that they are related to the symmetry elements in the same way. For instance, if $\mathbf{T}_s = 0$ for $s = 1, \ldots, 8$, then the allowed origins in *Pmmm* are the eight inversion centres.

To each functional form of the s.f. a set of permissible origins will correspond.

(*c*) A translation between permissible origins will be called a *permissible* or *allowed translation.* Trivial allowed translations correspond to the lattice periods or to their multiples. A change of origin by an allowed translation does not change the algebraic form of the s.f. Thus, according to (2.2.3.2), all origins allowed by a fixed functional form of the s.f. will be connected by translational vectors $\mathbf{X}_p$ such that

$$(\mathbf{R}_s - \mathbf{I})\mathbf{X}_p = \mathbf{V}, \quad s = 1, 2, \ldots, m, \qquad (2.2.3.3)$$

where $\mathbf{V}$ is a vector with zero or integer components.

In centred space groups, an origin translation corresponding to a centring vector $\mathbf{B}_v$ does not change the functional form of the s.f. Therefore all vectors $\mathbf{B}_v$ represent permissible translations. $\mathbf{X}_p$ will then be an allowed translation (Giacovazzo, 1974) not only when, as imposed by (2.2.3.3), the difference $\mathbf{T}'_s - \mathbf{T}_s$ is equal to one or more lattice units, but also when, for any s, the condition

$$(\mathbf{R}_s - \mathbf{I})\mathbf{X}_p = \mathbf{V} + \alpha\mathbf{B}_v, \quad s = 1, 2, \ldots, m; \quad \alpha = 0, 1 \qquad (2.2.3.4)$$

is satisfied.

We will call any set of cs. or ncs. space groups having the same allowed origin translations a Hauptman–Karle group (H–K group). The 94 ncs. primitive space groups, the 62 primitive cs. groups, the 44 ncs. centred space groups and the 30 cs. centred space groups can be collected into 13, 4, 14 and 5 H–K groups, respectively (Hauptman & Karle, 1953, 1956; Karle & Hauptman, 1961; Lessinger & Wondratschek, 1975). In Tables 2.2.3.1–2.2.3.4 the H–K groups are given together with the allowed origin translations.

(*d*) Let us consider a product of structure factors

$$F_{\mathbf{h}_1}^{A_1} \times F_{\mathbf{h}_2}^{A_2} \times \ldots \times F_{\mathbf{h}_n}^{A_n} = \prod_{j=1}^{n} F_{\mathbf{h}_j}^{A_j} = \exp\left(i\sum_{j=1}^{n} A_j\varphi_{\mathbf{h}_j}\right)\prod_{j=1}^{n} |F_{\mathbf{h}_j}|^{A_j}, \qquad (2.2.3.5)$$

A_j being integer numbers.

Copyright © 2010 International Union of Crystallography

The factor $\sum_{j=1}^{n} A_j\varphi_{\mathbf{h}_j}$ is the phase of the product (2.2.3.5). A *structure invariant* (s.i.) is a product (2.2.3.5) such that

$$\sum_{j=1}^{n} A_j\mathbf{h}_j = 0. \qquad (2.2.3.6)$$

Since $|F_{\mathbf{h}_j}|$ are usually known from experiment, it is often said that s.i.'s are combinations of phases

$$\sum_{j=1}^{n} A_j\varphi_{\mathbf{h}_j}, \qquad (2.2.3.7)$$

for which (2.2.3.6) holds.

F_0, $F_{\mathbf{h}}F_{-\mathbf{h}}$, $F_{\mathbf{h}}F_{\mathbf{k}}F_{\overline{\mathbf{h}+\mathbf{k}}}$, $F_{\mathbf{h}}F_{\mathbf{k}}F_{\mathbf{l}}F_{\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}}}$, $F_{\mathbf{h}}F_{\mathbf{k}}F_{\mathbf{l}}F_{\mathbf{p}}F_{\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{p}}}$ are examples of s.i.'s for $n = 1, 2, 3, 4, 5$.

The value of any s.i. does not change with an arbitrary shift of the space-group origin and thus it will depend on the crystal structure only.

(*e*) A *structure seminvariant* (s.s.) is a product of structure factors [or a combination of phases (2.2.3.7)] whose value is unchanged when the origin is moved by an allowed translation.

Let $\mathbf{X}_p$'s be the permissible origin translations of the space group. Then the product (2.2.3.5) [or the sum (2.2.3.7)] is an s.s., if, in accordance with (2.2.3.1),

$$\sum_{j=1}^{n} A_j(\mathbf{h}_j \cdot \mathbf{X}_p) = r, \quad p = 1, 2, \ldots \qquad (2.2.3.8)$$

where r is a positive integer, null or a negative integer.

Conditions (2.2.3.8) can be written in the following more useful form (Hauptman & Karle, 1953):

$$\sum_{j=1}^{n} A_j\mathbf{h}_{s_j} \equiv 0 \pmod{\boldsymbol{\omega}_s}, \qquad (2.2.3.9)$$

where $\mathbf{h}_{s_j}$ is the vector seminvariantly associated with the vector $\mathbf{h}_j$ and $\boldsymbol{\omega}_s$ is the seminvariant modulus. In Tables 2.2.3.1–2.2.3.4, the reflection $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$, the seminvariant modulus $\boldsymbol{\omega}_s$ and seminvariant phases are given for every H–K group.

The symbol of any group (*cf.* Giacovazzo, 1974) has the structure $\mathbf{h}_s L\boldsymbol{\omega}_s$, where L stands for the lattice symbol. This symbol is underlined if the space group is cs.

By definition, if the class of permissible origin has been chosen, that is to say, if the algebraic form of the symmetry operators has been fixed, then the value of an s.s. does not depend on the origin but on the crystal structure only.

(*f*) Suppose that we have chosen the symmetry operators $\mathbf{C}_s$ and thus fixed the functional form of the s.f.'s and the set of allowed origins. In order to describe the structure in direct space a unique reference origin must be fixed. Thus the phase-determining process must also require a unique permissible origin congruent to the values assigned to the phases. More specifically, at the beginning of the structure-determining process by direct methods we shall assign as many phases as necessary to define a unique origin among those allowed (and, as we shall see, possibly to fix the enantiomorph). From the theory developed so far it is obvious that arbitrary phases can be assigned to one or more s.f.'s if there is at least one allowed origin which, fixed as the origin of the unit cell, will give those phase values to the chosen reflections. The concept of *linear dependence* will help us to fix the origin.

(*g*) n phases $\varphi_{\mathbf{h}_j}$ are *linearly semidependent* (Hauptman & Karle, 1956) when the n vectors $\mathbf{h}_{s_j}$ seminvariantly associated with the $\mathbf{h}_j$ are linearly dependent modulo $\boldsymbol{\omega}_s$, $\boldsymbol{\omega}_s$ being the seminvariant modulus of the space group. In other words, when

$$\sum_{j=1}^{n} A_j\mathbf{h}_{s_j} \equiv 0 \pmod{\boldsymbol{\omega}_s}, \qquad A_q \not\equiv 0 \pmod{\boldsymbol{\omega}_s} \qquad (2.2.3.10)$$

is satisfied. The second condition means that at least one A_q exists that is not congruent to zero modulo each of the components of $\boldsymbol{\omega}_s$. If (2.2.3.10) is not satisfied for any n-set of integers A_j, the phases $\varphi_{\mathbf{h}_j}$ are *linearly semindependent.* If (2.2.3.10) is valid for $n = 1$ and $A = 1$, then $\mathbf{h}_1$ is said to be linearly semidependent and $\varphi_{\mathbf{h}_1}$ is an s.s. It may be concluded that a seminvariant phase is linearly semidependent, and, *vice versa*, that a phase linearly semidependent is an s.s. In Tables 2.2.3.1–2.2.3.4 the allowed variations (which are those due to the allowed origin translations) for the semindependent phases are given for every H–K group. If $\varphi_{\mathbf{h}_1}$ is linearly semindependent its value can be fixed arbitrarily because at least one origin compatible with the given value exists. Once $\varphi_{\mathbf{h}_1}$ is assigned, the necessary condition to be able to fix a second phase $\varphi_{\mathbf{h}_2}$ is that it should be linearly semindependent of $\varphi_{\mathbf{h}_1}$.

Similarly, the necessary condition to be able arbitrarily to assign a third phase $\varphi_{\mathbf{h}_3}$ is that it should be linearly semindependent from $\varphi_{\mathbf{h}_1}$ and $\varphi_{\mathbf{h}_2}$.

In general, the number of linearly semindependent phases is equal to the dimension of the seminvariant vector $\boldsymbol{\omega}_s$ (see Tables 2.2.3.1–2.2.3.4). The reader will easily verify in (h, k, l) P (2, 2, 2) that the three phases φ_{oee}, φ_{eoe}, φ_{eoo} define the origin (o indicates odd, e even).

(*h*) From the theory summarized so far it is clear that a number of semindependent phases $\varphi_{\mathbf{h}_j}$, equal to the dimension of the seminvariant vector $\boldsymbol{\omega}_s$, may be arbitrarily assigned in order to fix the origin. However, it is not always true that only one allowed origin compatible with the given phases exists. An additional condition is required such that only one permissible origin should lie at the intersection of the lattice planes corresponding to the origin-fixing reflections (or on the lattice plane $\mathbf{h}$ if one reflection is sufficient to define the origin). It may be shown that the condition is verified if the determinant formed with the vectors seminvariantly associated with the origin reflections, reduced modulo $\boldsymbol{\omega}_s$, has the value ± 1. In other words, such a determinant should be primitive modulo $\boldsymbol{\omega}_s$.

For example, in $P\bar{1}$ the three reflections

$$\mathbf{h}_1 = (345), \mathbf{h}_2 = (139), \mathbf{h}_3 = (784)$$

define the origin uniquely because

$$\begin{vmatrix} 3 & 4 & 5 \\ 1 & 3 & 9 \\ 7 & 8 & 4 \end{vmatrix} \xrightarrow{\text{reduced mod (2,2,2)}} \begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{vmatrix} = -1.$$

Furthermore, in $P4mm$ [$\mathbf{h}_s = (h + k, l)$, $\boldsymbol{\omega}_s = (2, 0)$]

$$\mathbf{h}_1 = (5, 2, 0), \quad \mathbf{h}_2 = (6, 2, 1)$$

define the origin uniquely since

$$\begin{vmatrix} 7 & 0 \\ 8 & 1 \end{vmatrix} \xrightarrow{\text{reduced mod (2,0)}} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

(*i*) If an s.s. or an s.i. has a general value φ for a given structure, it will have a value $-\varphi$ for the enantiomorph structure. If $\varphi = 0, \pi$ the s.s. has the same value for both enantiomorphs. Once the origin has been assigned, in ncs. space groups the sign of a given s.s. $\varphi \neq 0, \pi$ can be assigned to fix the enantiomorph. In practice it is often advisable to use an s.s. or an s.i. whose value is as near as possible to $\pm\pi/2$.

2.2.4. Normalized structure factors

2.2.4.1. Definition of normalized structure factor

The normalized structure factors E (see also Chapter 2.1) are calculated according to (Hauptman & Karle, 1953)

$$|E_{\mathbf{h}}|^2 = |F_{\mathbf{h}}|^2/\langle|F_{\mathbf{h}}|^2\rangle, \qquad (2.2.4.1)$$

Table 2.2.3.1. *Allowed origin translations, seminvariant moduli and phases for centrosymmetric primitive space groups*

	H–K group					
	$(h, k, l)\underline{P}(2, 2, 2)$		$(h+k, l)\underline{P}(2, 2)$		$(l)\underline{P}(2)$	$(h+k+l)\underline{P}(2)$
Space group	$P\bar{1}$	$Pmna$	$P\frac{4}{m}$	$P\frac{4}{n}mm$	$P\bar{3}$	$R\bar{3}$
	$P\frac{2}{m}$	$Pcca$	$P\frac{4_2}{m}$	$P\frac{4}{n}cc$	$P\bar{3}1m$	$R\bar{3}m$
	$P\frac{2_1}{m}$	$Pbam$	$P\frac{4}{n}$	$P\frac{4_2}{m}mc$	$P\bar{3}1c$	$R\bar{3}c$
	$P\frac{2}{c}$	$Pccn$	$P\frac{4_2}{n}$	$P\frac{4_2}{m}cm$	$P\bar{3}m1$	$Pm\bar{3}$
	$P\frac{2_1}{c}$	$Pbcm$	$P\frac{4}{m}mm$	$P\frac{4_2}{n}bc$	$P\bar{3}c1$	$Pn\bar{3}$
	$Pmmm$	$Pnnm$	$P\frac{4}{m}cc$	$P\frac{4_2}{n}nm$	$P\frac{6}{m}$	$Pa\bar{3}$
	$Pnnn$	$Pmmn$	$P\frac{4}{n}bm$	$P\frac{4_2}{m}bc$	$P\frac{6_3}{m}$	$Pm\bar{3}m$
	$Pccm$	$Pbcn$	$P\frac{4}{n}nc$	$P\frac{4_2}{m}nm$	$P\frac{6}{m}mm$	$Pn\bar{3}n$
	$Pban$	$Pbca$	$P\frac{4}{m}bm$	$P\frac{4_2}{n}mc$	$P\frac{6}{m}cc$	$Pm\bar{3}n$
	$Pmma$	$Pnma$	$P\frac{4}{m}nc$	$P\frac{4_2}{n}cm$	$P\frac{6_3}{m}cm$	$Pn\bar{3}m$
	$Pnna$				$P\frac{6_3}{m}mc$	
Allowed origin translations	$(0, 0, 0)$; $(\frac{1}{2}, 0, 0)$; $(0, \frac{1}{2}, 0)$; $(0, 0, \frac{1}{2})$;	$(0, \frac{1}{2}, \frac{1}{2})$ $(\frac{1}{2}, 0, \frac{1}{2})$ $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(0, 0, 0)$ $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$		$(0, 0, 0)$ $(0, 0, \frac{1}{2})$	$(0, 0, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(h, k, l)		$(h+k, l)$		(l)	$(h+k+l)$
Seminvariant modulus $\boldsymbol{\omega}_s$	(2, 2, 2)		(2, 2)		(2)	(2)
Seminvariant phases	φ_{eee}		φ_{eee}; φ_{ooe}		φ_{eee}; φ_{eoe} φ_{oee}; φ_{ooe}	φ_{eee}; φ_{ooe} φ_{oeo}; φ_{eoo}
Number of semindependent phases to be specified	3		2		1	1

where $|F_{\mathbf{h}}|^2$ is the squared observed structure-factor magnitude on the absolute scale and $\langle|F_{\mathbf{h}}|^2\rangle$ is the expected value of $|F_{\mathbf{h}}|^2$.

$\langle|F_{\mathbf{h}}|^2\rangle$ depends on the available *a priori* information. Often, but not always, this may be considered as a combination of several typical situations. We mention:

(*a*) No structural information. The atomic positions are considered random variables. Then

$$\langle|F_{\mathbf{h}}|^2\rangle = \varepsilon_{\mathbf{h}} \sum_{j=1}^{N} f_j^2 = \varepsilon_{\mathbf{h}} \sum\nolimits_N$$

so that

$$E_{\mathbf{h}} = \frac{F_{\mathbf{h}}}{(\varepsilon_{\mathbf{h}} \sum\nolimits_N)^{1/2}}. \qquad (2.2.4.2)$$

$\varepsilon_{\mathbf{h}}$ takes account of the effect of space-group symmetry (see Chapter 2.1).

(*b*) P atomic groups having a known configuration but with unknown orientation and position (Main, 1976). Then a certain number of interatomic distances $r_{j_1 j_2}$ are known and

$$\langle|F_{\mathbf{h}}|^2\rangle = \varepsilon_{\mathbf{h}}\left(\sum\nolimits_N + \sum_{i=1}^{P} \sum_{j_1 \neq j_2 = 1}^{M_i} f_{j_1} f_{j_2} \frac{\sin 2\pi q r_{j_1 j_2}}{2\pi q r_{j_1 j_2}}\right),$$

where M_i is the number of atoms in the ith molecular fragment and $q = |\mathbf{h}|$.

(*c*) P atomic groups with a known configuration, correctly oriented, but with unknown position (Main, 1976). Then a certain group of interatomic vectors $\mathbf{r}_{j_1 j_2}$ is fixed and

$$\langle|F_{\mathbf{h}}|^2\rangle = \varepsilon_{\mathbf{h}}\left(\sum\nolimits_N + \sum_{i=1}^{P} \sum_{j_1 \neq j_2 = 1}^{M_i} f_{j_1} f_{j_2} \exp 2\pi i \mathbf{h} \cdot \mathbf{r}_{j_1 j_2}\right).$$

The above formula has been derived on the assumption that primitive positional random variables are uniformly distributed over the unit cell. Such an assumption may be considered unfavourable (Giacovazzo, 1988) in space groups for which the allowed shifts of origin, consistent with the chosen algebraic form for the symmetry operators $\mathbf{C}_s$, are arbitrary displacements along any polar axes. Thanks to the indeterminacy in the choice of origin, the first of the shifts $\boldsymbol{\tau}_i$ (to be applied to the ith fragment in order to translate atoms in the correct positions) may be restricted to a region which is smaller than the unit cell (*e.g.* in $P2$ we are free to specify the origin along the diad axis by restricting $\boldsymbol{\tau}_1$ to the family of vectors $\{\boldsymbol{\tau}_1\}$ of type $[x0z]$). The practical consequence is that $\langle|F_{\mathbf{h}}|^2\rangle$ is significantly modified in polar space groups if $\mathbf{h}$ satisfies

$$\mathbf{h} \cdot \boldsymbol{\tau}_1 = 0,$$

where $\boldsymbol{\tau}_1$ belongs to the family of restricted vectors $\{\boldsymbol{\tau}_1\}$.

(*d*) Atomic groups correctly positioned. Then (Main, 1976; Giacovazzo, 1983*a*)

$$\langle|F_{\mathbf{h}}|^2\rangle = |F_{\mathrm{p},\mathbf{h}}|^2 + \varepsilon_{\mathbf{h}}\sum_q,$$

where $F_{\mathrm{p},\mathbf{h}}$ is the structure factor of the partial known structure and q are the atoms with unknown positions.

(*e*) A pseudotranslational symmetry is present. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots$ be the pseudotranslation vectors of order $n_1, n_2, n_3, \ldots$, respectively. Furthermore, let p be the number of atoms (symmetry equivalents included) whose positions are related by pseudotranslational symmetry and q the number of atoms (symmetry equivalents included) whose positions are not related by any pseudotranslation. Then (Cascarano *et al.*, 1985*a*,*b*)

$$\langle|F_{\mathbf{h}}|^2\rangle = \varepsilon_{\mathbf{h}}(\zeta_{\mathbf{h}}\sum_p + \sum_q),$$

where

$$\zeta_{\mathbf{h}} = \frac{(n_1 n_2 n_3 \ldots)\gamma_{\mathbf{h}}}{m}$$

and $\gamma_{\mathbf{h}}$ is the number of times for which algebraic congruences

$$\mathbf{h}\cdot\mathbf{R}_s\mathbf{u}_i \equiv 0 \pmod 1 \quad \text{for } i = 1, 2, 3, \ldots$$

are simultaneously satisfied when s varies from 1 to m. If $\gamma_{\mathbf{h}} = 0$ then $F_{\mathbf{h}}$ is said to be a *superstructure reflection*, otherwise it is a *substructure reflection.*

Often substructures are not ideal: *e.g.* atoms related by pseudotranslational symmetry are ideally located but of different type (replacive deviations from ideality); or they are equal but not ideally located (displacive deviations); or a combination of the two situations occurs. In these cases a correlation exists between the substructure and the superstructure. It has been shown (Mackay, 1953; Cascarano *et al.*, 1988*a*) that the scattering power of the substructural part may be estimated *via* a statistical analysis of diffraction data for ideal pseudotranslational symmetry or for displacive deviations from it, while it is not estimable in the case of replacive deviations.

2.2.4.2. Definition of quasi-normalized structure factor

When probability theory is not used, the *quasi-normalized structure factors* $\mathscr{E}_{\mathbf{h}}$ and the *unitary structure factors* $U_{\mathbf{h}}$ are often used. $\mathscr{E}_{\mathbf{h}}$ and $U_{\mathbf{h}}$ are defined according to

$$|\mathscr{E}_{\mathbf{h}}|^2 = \varepsilon_{\mathbf{h}}|E_{\mathbf{h}}|^2$$

$$U_{\mathbf{h}} = F_{\mathbf{h}}\Big/\left(\sum_{j=1}^{N} f_j\right).$$

Since $\sum_{j=1}^{N} f_j$ is the largest possible value for $F_{\mathbf{h}}$, $U_{\mathbf{h}}$ represents the fraction of $F_{\mathbf{h}}$ with respect to its largest possible value. Therefore

Table 2.2.3.2. *Allowed origin translations, seminvariant moduli and phases for noncentrosymmetric primitive space groups*

	H–K group					
	$(h, k, l)P(0, 0, 0)$	$(h, k, l)P(2, 0, 2)$	$(h, k, l)P(0, 2, 0)$	$(h, k, l)P(2, 2, 2)$	$(h, k, l)P(2, 2, 0)$	$(h + k, l)P(2, 0)$
Space group	*P*1	*P*2 $P2_1$	*Pm* *Pc*	*P*222 $P222_1$ $P2_12_12$ $P2_12_12_1$	*Pmm*2 $Pmc2_1$ *Pcc*2 *Pma*2 $Pca2_1$ *Pnc*2 $Pmn2_1$ *Pba*2 $Pna2_1$ *Pnn*2	*P*4 $P4_1$ $P4_2$ $P4_3$ *P*4*mm* *P*4*bm* $P4_2cm$ $P4_2nm$ *P*4*cc* *P*4*nc* $P4_2mc$ $P4_2bc$
Allowed origin translations	(x, y, z)	$(0, y, 0)$ $(0, y, \frac{1}{2})$ $(\frac{1}{2}, y, 0)$ $(\frac{1}{2}, y, \frac{1}{2})$	$(x, 0, z)$ $(x, \frac{1}{2}, z)$	$(0, 0, 0)$ $(\frac{1}{2}, 0, 0)$ $(0, \frac{1}{2}, 0)$ $(0, 0, \frac{1}{2})$ $(0, \frac{1}{2}, \frac{1}{2})$ $(\frac{1}{2}, 0, \frac{1}{2})$ $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(0, 0, z)$ $(0, \frac{1}{2}, z)$ $(\frac{1}{2}, 0, z)$ $(\frac{1}{2}, \frac{1}{2}, z)$	$(0, 0, z)$ $(\frac{1}{2}, \frac{1}{2}, z)$
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(h, k, l)	(h, k, l)	(h, k, l)	(h, k, l)	(h, k, l)	$(h + k, l)$
Seminvariant modulus $\boldsymbol{\omega}_s$	(0, 0, 0)	(2, 0, 2)	(0, 2, 0)	(2, 2, 2)	(2, 2, 0)	(2, 0)
Seminvariant phases	φ_{000}	φ_{e0e}	$\varphi_{0\mathrm{e}0}$	φ_{eee}	$\varphi_{\mathrm{ee}0}$	$\varphi_{\mathrm{ee}0}$ $\varphi_{\mathrm{oo}0}$
Allowed variations for the semindependent phases	$\|\infty\|$	$\|\infty\|$, $\|2\|$ if $k = 0$	$\|\infty\|$, $\|2\|$ if $h = l = 0$	$\|2\|$	$\|\infty\|$, $\|2\|$ if $l = 0$	$\|\infty\|$, $\|2\|$ if $l = 0$
Number of semindependent phases to be specified	3	3	3	3	3	2

$$0 \leq |U_{\mathbf{h}}| \leq 1.$$

If atoms are equal, then $U_{\mathbf{h}} = \mathscr{E}_{\mathbf{h}}/N^{1/2}$.

2.2.4.3. The calculation of normalized structure factors

N.s.f.'s cannot be calculated by applying (2.2.4.1) to observed s.f.'s because: (*a*) the observed magnitudes $I_{\mathbf{h}}$ (already corrected for Lp factor, absorption, ...) are on a relative scale; (*b*) $\langle|F_{\mathbf{h}}|^2\rangle$ cannot be calculated without having estimated the vibrational motion of the atoms.

This is usually obtained by the well known Wilson plot (Wilson, 1942), according to which observed data are divided into ranges of $s^2 = \sin^2\theta/\lambda^2$ and averages of intensity $\langle I_{\mathbf{h}}\rangle$ are taken in each shell. Reflection multiplicities and other effects of space-group symmetry on intensities must be taken into account when such averages are calculated. The shells are symmetrically overlapped in order to reduce statistical fluctuations and are restricted so that the number of reflections in each shell is reasonably large. For each shell

$$K\langle I\rangle = \langle|F|^2\rangle = \langle|F^{\rm o}|^2\rangle \exp(-2Bs^2) \qquad (2.2.4.3)$$

should be obtained, where K is the scale factor needed to place X-ray intensities on the absolute scale, B is the overall thermal parameter and $\langle|F^{\rm o}|^2\rangle$ is the expected value of $|F|^2$ in which it is assumed that all the atoms are at rest. $\langle|F^{\rm o}|^2\rangle$ depends upon the structural information that is available (see Section 2.2.4.1 for some examples).

Equation (2.2.4.3) may be rewritten as

$$\ln\left\{\frac{\langle I\rangle}{\langle|F^{\rm o}|^2\rangle}\right\} = -\ln K - 2Bs^2,$$

which plotted at various s^2 should be a straight line of which the slope ($2B$) and intercept ($\ln K$) on the logarithmic axis can be obtained by applying a linear least-squares procedure.

Very often molecular geometries produce perceptible departures from linearity in the logarithmic Wilson plot. However, the more extensive the available *a priori* information on the structure is, the closer, on the average, are the Wilson-plot curves to their least-squares straight lines.

Accurate estimates of B and K require good strategies (Rogers & Wilson, 1953) for:

(1) treatment of weak measured data. If weak data are set to zero, there will be bias in the statistics. Methods are, however, available (French & Wilson, 1978) that provide an *a posteriori* estimate of weak (even negative) intensities by means of Bayesian statistics.

Table 2.2.3.2 (*cont.*)

$(h+k, l)P(2, 2)$	$(h-k, l)P(3, 0)$	$(2h+4k+3l)P(6)$	$(l)P(0)$	$(l)P(2)$	$(h+k+l)P(0)$	$(h+k+l)P(2)$
$P\bar{4}$	$P3$	$P312$	$P31m$	$P321$	$R3$	$R32$
$P422$	$P3_1$	$P3_112$	$P31c$	$P3_121$	$R3m$	$P23$
$P42_12$	$P3_2$	$P3_212$	$P6$	$P3_221$	$R3c$	$P2_13$
$P4_122$	$P3m1$	$P6$	$P6_1$	$P622$		$P432$
$P4_12_12$	$P3c1$	$P\bar{6}m2$	$P6_5$	$P6_122$		$P4_232$
$P4_222$		$P\bar{6}c2$	$P6_4$	$P6_522$		$P4_332$
$P4_22_12$			$P6_3$	$P6_222$		$P4_132$
$P4_322$			$P6_2$	$P6_422$		$P\bar{4}3m$
$P4_32_12$			$P6mm$	$P6_322$		$P\bar{4}3n$
$P\bar{4}2m$			$P6cc$	$P\bar{6}2m$		
$P\bar{4}2c$			$P6_3cm$	$P\bar{6}2c$		
$P\bar{4}2_1m$			$P6_3mc$			
$P\bar{4}2_1c$						
$P\bar{4}m2$						
$P\bar{4}c2$						
$P\bar{4}b2$						
$P\bar{4}n2$						
(0, 0, 0) (0, 0, $\frac{1}{2}$) ($\frac{1}{2}$, $\frac{1}{2}$, 0) ($\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$)	(0, 0, z) ($\frac{1}{3}$, $\frac{2}{3}$, z) ($\frac{2}{3}$, $\frac{1}{3}$, z)	(0, 0, 0) (0, 0, $\frac{1}{2}$) ($\frac{1}{3}$, $\frac{2}{3}$, 0) ($\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{2}$) ($\frac{2}{3}$, $\frac{1}{3}$, 0) ($\frac{2}{3}$, $\frac{1}{3}$, $\frac{1}{2}$)	(0, 0, z)	(0, 0, 0) (0, 0, $\frac{1}{2}$)	(x, x, x)	(0, 0, 0) ($\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$)
$(h+k, l)$	$(h-k, l)$	$(2h+4k+3l)$	(l)	(l)	$(h+k+l)$	$(h+k+l)$
(2, 2)	(3, 0)	(6)	(0)	(2)	(0)	(2)
$\varphi_{\rm eee}$ $\varphi_{\rm ooe}$	φ_{hk0} if $h-k=0$ (mod 3)	φ_{hkl} if $2h+4k+3l=0$ (mod 6)	φ_{hk0}	φ_{hke}	$\varphi_{h,k,\bar{h}+\bar{k}}$	$\varphi_{\rm eee}$; $\varphi_{\rm ooe}$ $\varphi_{\rm oee}$; $\varphi_{\rm ooe}$
$\|2\|$	$\|\infty\|$, $\|3\|$ if $l=0$	$\|2\|$ if $h \equiv k$ (mod 3) $\|3\|$ if $l \equiv 0$ (mod 2)	$\|\infty\|$	$\|2\|$	$\|\infty\|$	$\|2\|$
2	2	1	1	1	1	1

Table 2.2.3.3. *Allowed origin translations, seminvariant moduli and phases for centrosymmetric non-primitive space groups*

	H–K group				
	$(h, l)\underline{C}(2, 2)$	$(k, l)\underline{I}(2, 2)$	$(h+k+l)\underline{F}(2)$	$(l)\underline{I}(2)$	$\underline{I}$
Space groups	$C\frac{2}{m}$ $C\frac{2}{c}$ $Cmcm$ $Cmca$ $Cmmm$ $Cccm$ $Cmma$ $Ccca$	$Immm$ $Ibam$ $Ibca$ $Imma$	$Fmmm$ $Fddd$ $Fm\bar{3}$ $Fd\bar{3}$ $Fm\bar{3}m$ $Fm\bar{3}c$ $Fd\bar{3}m$ $Fd\bar{3}c$	$I\frac{4}{m}$ $I\frac{4_1}{a}$ $I\frac{4}{m}mm$ $I\frac{4}{m}cm$ $I\frac{4_1}{a}md$ $I\frac{4_1}{a}cd$	$Im\bar{3}$ $Ia\bar{3}$ $Im\bar{3}m$ $Ia\bar{3}d$
Allowed origin translations	(0, 0, 0) $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, 0, 0)$ $(\frac{1}{2}, 0, \frac{1}{2})$	(0, 0, 0) $(0, 0, \frac{1}{2})$ $(0, \frac{1}{2}, 0)$ $(\frac{1}{2}, 0, 0)$	(0, 0, 0) $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	(0, 0, 0) $(0, 0, \frac{1}{2})$	(0, 0, 0)
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(h, l)	(k, l)	$(h+k+l)$	(l)	(h, k, l)
Seminvariant modulus $\boldsymbol{\omega}_s$	(2, 2)	(2, 2)	(2)	(2)	(1, 1, 1)
Seminvariant phases	φ_{eee}	φ_{eee}	φ_{eee}	φ_{eoe}; φ_{eee} φ_{ooe}; φ_{oee}	All
Number of semindependent phases to be specified	2	2	1	1	0

(2) treatment of missing weak data (Rogers *et al.*, 1955; Vicković & Viterbo, 1979). All unobserved reflections may assume

$$\mu = |F_{o\,\min}|^2/3 \text{ for cs. space groups}$$
$$\mu = |F_{o\,\min}|^2/2 \text{ for ncs. space groups,}$$

where the subscript 'o min' refers to the minimum observed intensity.

Once K and B have been estimated, $E_{\mathbf{h}}$ values can be obtained from experimental data by

$$|E_{\mathbf{h}}|^2 = \frac{KI_{\mathbf{h}}}{\langle |F_{\mathbf{h}}^{o}|^2 \rangle \exp(-2Bs^2)},$$

Table 2.2.3.4. *Allowed origin translations, seminvariant moduli and phases for noncentrosymmetric non-primitive space groups*

	H–K group					
	$(k, l)C(0, 2)$	$(h, l)C(0, 0)$	$(h, l)C(2, 0)$	$(h, l)C(2, 2)$	$(h, l)A(2, 0)$	$(h, l)I(2, 0)$
Space group	$C2$	Cm Cc	$Cmm2$ $Cmc2_1$ $Ccc2$	$C222$ $C222_1$	$Amm2$ $Abm2$ $Ama2$ $Aba2$	$Imm2$ $Iba2$ $Ima2$
Allowed origin translations	(0, y, 0) $(0, y, \frac{1}{2})$	(x, 0, z)	(0, 0, z) $(\frac{1}{2}, 0, z)$	(0, 0, 0) $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, 0, 0)$ $(\frac{1}{2}, 0, \frac{1}{2})$	(0, 0, z) $(\frac{1}{2}, 0, z)$	(0, 0, z) $(\frac{1}{2}, 0, z)$
Vector $\mathbf{h}_s$ seminvariantly associated with $\mathbf{h} = (h, k, l)$	(k, l)	(h, l)	(h, l)	(h, l)	(h, l)	(h, l)
Seminvariant modulus $\boldsymbol{\omega}_s$	(0, 2)	(0, 0)	(2, 0)	(2, 2)	(2, 0)	(2, 0)
Seminvariant phases	φ_{e0e}	φ_{0e0}	φ_{ee0}	φ_{eee}	φ_{ee0}	φ_{ee0}
Allowed variations for the semindependent phases	$\|\infty\|$, $\|2\|$ if $k = 0$	$\|\infty\|$	$\|\infty\|$, $\|2\|$ if $l = 0$	$\|2\|$	$\|\infty\|$, $\|2\|$ if $l = 0$	$\|\infty\|$, $\|2\|$ if $l = 0$
Number of semindependent phases to be specified	2	2	2	2	2	2

where $\langle|F_{\mathbf{h}}^{\mathrm{o}}|^2\rangle$ is the expected value of $|F_{\mathbf{h}}^{\mathrm{o}}|^2$ for the reflection **h** on the basis of the available *a priori* information.

2.2.4.4. Probability distributions of normalized structure factors

Under some fairly general assumptions (see Chapter 2.1) probability distribution functions for the variable $|E|$ for cs. and ncs. structures are (see Fig. 2.2.4.1)

$$ {}_{\bar{1}}P(|E|)\,\mathrm{d}|E| = \sqrt{\frac{2}{\pi}}\exp\left(-\frac{E^2}{2}\right)\mathrm{d}|E| \qquad (2.2.4.4)$$

and

$$ {}_{1}P(|E|)\,\mathrm{d}|E| = 2|E|\exp(-|E|^2)\,\mathrm{d}|E|, \qquad (2.2.4.5)$$

respectively. Corresponding cumulative functions are (see Fig. 2.2.4.2)

$$ {}_{\bar{1}}N(|E|) = \sqrt{\frac{2}{\pi}}\int_0^{|E|}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t = \mathrm{erf}\left(\frac{|E|}{\sqrt{2}}\right),$$

$$ {}_{1}N(|E|) = \int_0^{|E|} 2t\exp(-t^2)\,\mathrm{d}t = 1-\exp(-|E|^2).$$

Some moments of the distributions (2.2.4.4) and (2.2.4.5) are listed in Table 2.2.4.1. In the absence of other indications for a given crystal structure, a cs. or an ncs. space group will be preferred according to whether the statistical tests yield values closer to column 2 or to column 3 of Table 2.2.4.1.

For further details about the distribution of intensities see Chapter 2.1.

2.2.5. Phase-determining formulae

From the earliest periods of X-ray structure analysis several authors (Ott, 1927; Banerjee, 1933; Avrami, 1938) have tried to determine atomic positions directly from diffraction intensities.

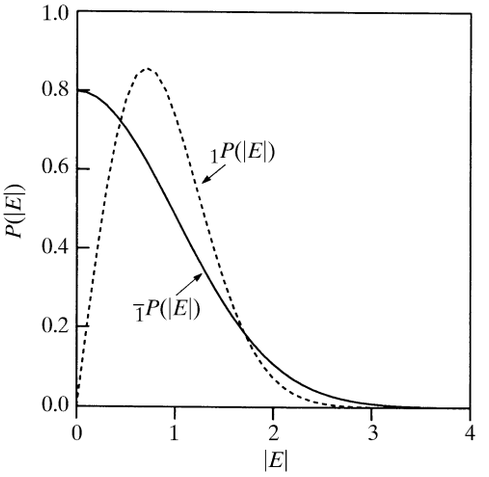


Fig. 2.2.4.1. Probability density functions for cs. and ncs. crystals.

Significant developments are the derivation of inequalities and the introduction of probabilistic techniques *via* the use of joint probability distribution methods (Hauptman & Karle, 1953).

2.2.5.1. Inequalities among structure factors

An extensive system of inequalities exists for the coefficients of a Fourier series which represents a positive function. This can restrict the allowed values for the phases of the s.f.'s in terms of measured structure-factor magnitudes. Harker & Kasper (1948) derived two types of inequalities:

Type 1. A modulus is bound by a combination of structure factors:

$$ |U_{\mathbf{h}}|^2 \le \frac{1}{m}\sum_{s=1}^{m} a_s(-\mathbf{h})U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)}, \qquad (2.2.5.1)$$

where m is the order of the point group and $a_s(-\mathbf{h}) = \exp(-2\pi i\mathbf{h}\cdot\mathbf{T}_s)$.

Table 2.2.3.4 (*cont.*)

$(h, l)I(2, 2)$	$(h+k+l)F(2)$	$(h+k+l)F(4)$	$(l)I(0)$	$(l)I(2)$	$(2k-l)I(4)$	$(l)F(0)$	I
$I222$ $I2_12_12_1$	$F432$ $F4_132$	$F222$ $F23$ $F\bar{4}3m$ $F\bar{4}3c$	$I4$ $I4_1$ $I4mm$ $I4cm$ $I4_1md$ $I4_1cd$	$I422$ $I4_122$ $I\bar{4}2m$ $I\bar{4}2d$	$I\bar{4}$ $I\bar{4}m2$ $I\bar{4}c2$	$Fmm2$ $Fdd2$	$I23$ $I2_13$ $I432$ $I4_132$ $I\bar{4}3m$ $I\bar{4}3d$
(0, 0, 0) $(0, 0, \frac{1}{2})$ $(0, \frac{1}{2}, 0)$ $(\frac{1}{2}, 0, 0)$	(0, 0, 0) $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	(0, 0, 0) $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ $(\frac{3}{4}, \frac{3}{4}, \frac{3}{4})$	$(0, 0, z)$	(0, 0, 0) $(0, 0, \frac{1}{2})$	(0, 0, 0) $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, 0, \frac{3}{4})$ $(\frac{1}{2}, 0, \frac{1}{4})$	$(0, 0, z)$	(0, 0, 0)
(h, l)	$(h+k+l)$	$(h+k+l)$	(l)	(l)	$(2k-l)$	(l)	(h, k, l)
(2, 2)	(2)	(4)	(0)	(2)	(4)	(0)	(1, 1, 1)
φ_{eee}	φ_{eee}	φ_{hkl} with $h+k+l \equiv 0$ (mod 4)	φ_{hk0}	φ_{hke}	φ_{hkl} with $(2k-l) \equiv 0$ (mod 4)	φ_{hk0}	All
$\|2\|$	$\|2\|$	$\|2\|$ if $h+k+l \equiv 0$ (mod 2) $\|4\|$ if $h+k+l \equiv 1$ (mod 2)	$\|\infty\|$	$\|2\|$	$\|2\|$ if $h+k+l \equiv 0$ (mod 2) $\|4\|$ if $2k-l \equiv 1$ (mod 2)	$\|\infty\|$	All
2	1	1	1	1	1	1	0

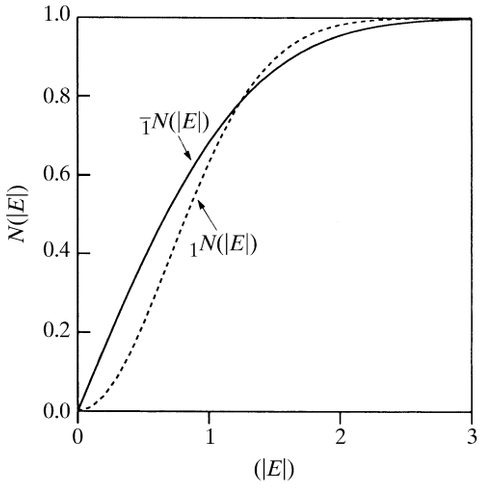


Fig. 2.2.4.2. Cumulative distribution functions for cs. and ncs. crystals.

Applied to low-order space groups, (2.2.5.1) gives

$$\begin{aligned} P1: &\quad |U_{h,k,l}|^2 \le 1 \\ P\bar{1}: &\quad U^2_{h,k,l} \le 0.5 + 0.5U_{2h,2k,2l} \\ P2_1: &\quad |U_{h,k,l}|^2 \le 0.5 + 0.5(-1)^k U_{2h,0,2l}. \end{aligned}$$

The meaning of each inequality is easily understandable: in $P\bar{1}$, for example, $U_{2h,2k,2l}$ must be positive if $|U_{h,k,l}|$ is large enough.

Type 2. The modulus of the sum or of the difference of two structure factors is bound by a combination of structure factors:

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \le \frac{1}{m}\left\{\sum_{s=1}^{m} a_s(-\mathbf{h})U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)} + \sum_{s=1}^{m} a_s(-\mathbf{h}')U_{\mathbf{h}'(\mathbf{I}-\mathbf{R}_s)} \pm 2\mathscr{R}e\left[\sum_{s=1}^{m} a_s(-\mathbf{h}')U_{\mathbf{h}-\mathbf{h}'\mathbf{R}_s}\right]\right\} \quad (2.2.5.2)$$

where $\mathscr{R}e$ stands for 'real part of'. Equation (2.2.5.2) applied to $P1$ gives

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \le 2 \pm 2|U_{\mathbf{h}-\mathbf{h}'}|\cos\varphi_{\mathbf{h}-\mathbf{h}'}.$$

A variant of (2.2.5.2) valid for cs. space groups is

$$(U_{\mathbf{h}} \pm U_{\mathbf{h}'})^2 \le (1 \pm U_{\mathbf{h}+\mathbf{h}'})(1 \pm U_{\mathbf{h}-\mathbf{h}'}).$$

After Harker & Kasper's contributions, several other inequalities were discovered (Gillis, 1948; Goedkoop, 1950; Okaya & Nitta, 1952; de Wolff & Bouman, 1954; Bouman, 1956; Oda *et al.*, 1961). The most general are the Karle–Hauptman inequalities (Karle & Hauptman, 1950):

$$D_m = \begin{vmatrix} U_0 & U_{-\mathbf{h}_1} & U_{-\mathbf{h}_2} & \cdots & U_{-\mathbf{h}_n} \\ U_{\mathbf{h}_1} & U_0 & U_{\mathbf{h}_1-\mathbf{h}_2} & \cdots & U_{\mathbf{h}_1-\mathbf{h}_n} \\ U_{\mathbf{h}_2} & U_{\mathbf{h}_2-\mathbf{h}_1} & U_0 & \cdots & U_{\mathbf{h}_2-\mathbf{h}_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{\mathbf{h}_n} & U_{\mathbf{h}_n-\mathbf{h}_1} & U_{\mathbf{h}_n-\mathbf{h}_2} & \cdots & U_0 \end{vmatrix} \ge 0. \quad (2.2.5.3)$$

The determinant can be of any order but the leading column (or row) must consist of U's with different indices, although, within the column, symmetry-related U's may occur. For $n = 2$ and $\mathbf{h}_2 = 2\mathbf{h}_1 = 2\mathbf{h}$, equation (2.2.5.3) reduces to

$$D_3 = \begin{vmatrix} U_0 & U_{-\mathbf{h}} & U_{-2\mathbf{h}} \\ U_{\mathbf{h}} & U_0 & U_{-\mathbf{h}} \\ U_{2\mathbf{h}} & U_{\mathbf{h}} & U_0 \end{vmatrix} \ge 0,$$

which, for cs. structures, gives the Harker & Kasper inequality

Table 2.2.4.1. *Moments of the distributions (2.2.4.4) and (2.2.4.5)*

$R(E_s)$ is the percentage of n.s.f.'s with amplitude greater than the threshold E_s.

Criterion	Centrosymmetric distribution	Noncentrosymmetric distribution
$\langle\|E\|\rangle$	0.798	0.886
$\langle\|E\|^2\rangle$	1.000	1.000
$\langle\|E\|^3\rangle$	1.596	1.329
$\langle\|E\|^4\rangle$	3.000	2.000
$\langle\|E\|^5\rangle$	6.383	3.323
$\langle\|E\|^6\rangle$	15.000	6.000
$\langle\|E^2-1\|\rangle$	0.968	0.736
$\langle(E^2-1)^2\rangle$	2.000	1.000
$\langle(E^2-1)^3\rangle$	8.000	2.000
$\langle\|E^2-1\|^3\rangle$	8.691	2.415
$R(1)$	0.320	0.368
$R(2)$	0.050	0.018
$R(3)$	0.003	0.0001

$$U^2_{\mathbf{h}} \le 0.5 + 0.5U_{2\mathbf{h}}.$$

For $m = 3$, equation (2.2.5.3) becomes

$$D_3 = \begin{vmatrix} U_0 & U_{-\mathbf{h}} & U_{-\mathbf{k}} \\ U_{\mathbf{h}} & U_0 & U_{\mathbf{h}-\mathbf{k}} \\ U_{\mathbf{k}} & U_{\mathbf{k}-\mathbf{h}} & U_0 \end{vmatrix} \ge 0,$$

from which

$$1 - |U_{\mathbf{h}}|^2 - |U_{\mathbf{k}}|^2 - |U_{\mathbf{h}-\mathbf{k}}|^2 + 2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|\cos\alpha_{\mathbf{h},\mathbf{k}} \ge 0, \quad (2.2.5.4)$$

where

$$\alpha_{\mathbf{h},\mathbf{k}} = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}.$$

If the moduli $|U_{\mathbf{h}}|$, $|U_{\mathbf{k}}|$, $|U_{\mathbf{h}-\mathbf{k}}|$ are large enough, (2.2.5.4) is not satisfied for all values of $\alpha_{\mathbf{h},\mathbf{k}}$. In cs. structures the eventual check that one of the two values of $\alpha_{\mathbf{h},\mathbf{k}}$ does not satisfy (2.2.5.4) brings about the unambiguous identification of the sign of the product $U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}$.

It was observed (Gillis, 1948) that 'there was a number of cases in which both signs satisfied the inequality, one of them by a comfortable margin and the other by only a relatively small margin. In almost all such cases it was the former sign which was the correct one. That suggests that the method may have some power in reserve in the sense that there are still fundamentally stronger inequalities to be discovered'. Today we identify this power in reserve in the use of probability theory.

2.2.5.2. Probabilistic phase relationships for structure invariants

For any space group (see Section 2.2.3) there are linear combinations of phases with cosines that are, in principle, fixed by the $|E|$ magnitudes alone (s.i.'s) or by the $|E|$ values and the trigonometric form of the structure factor (s.s.'s). This result greatly stimulated the calculation of conditional distribution functions

$$P(\Phi|\{R\}), \quad (2.2.5.5)$$

where $R_{\mathbf{h}} = |E_{\mathbf{h}}|$, $\Phi = \sum A_i\varphi_{\mathbf{h}_i}$ is an s.i. or an s.s. and $\{R\}$ is a suitable set of diffraction magnitudes. The method was first proposed by Hauptman & Karle (1953) and was developed further by several authors (Bertaut, 1955*a*,*b*, 1960; Klug, 1958; Naya *et al.*, 1964, 1965; Giacovazzo, 1980*a*). From a probabilistic point of view the crystallographic problem is clear: the joint distribution $P(E_{\mathbf{h}_1}, \ldots, E_{\mathbf{h}_n})$, from which the conditional distributions (2.2.5.5) can be derived, involves a number of normalized structure factors each of which is a linear sum of random vari-

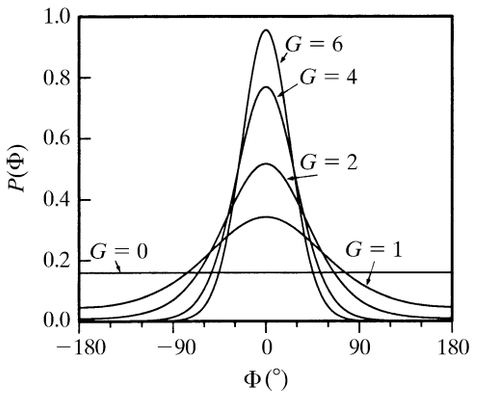


Fig. 2.2.5.1. Curves of (2.2.5.6) for some values of $G = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|$.

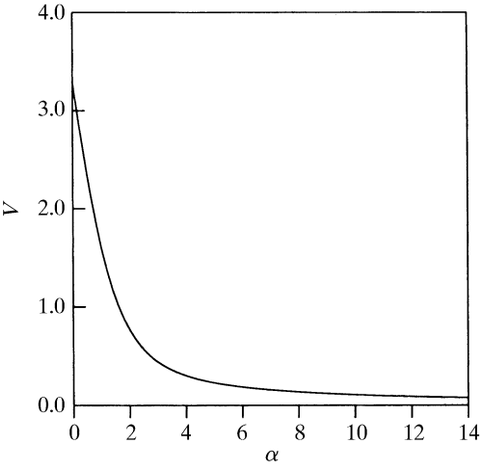


Fig. 2.2.5.2. Variance (in square radians) as a function of α.

ables (the atomic contributions to the structure factors). So, for the probabilistic interpretation of the phase problem, the atomic positions and the reciprocal vectors may be considered as random variables. A further problem is that of identifying, for a given Φ, a suitable set of magnitudes $|E|$ on which Φ primarily depends. The formulation of the *nested neighbourhood principle* first (Hauptman, 1975) fixed the idea of defining a sequence of sets of reflections each contained in the succeeding one and having the property that any s.i. or s.s. may be estimated *via* the magnitudes constituting the various neighbourhoods. A subsequent more general theory, the *representation method* (Giacovazzo, 1977*a*, 1980*b*), arranges for any Φ the set of intensities in a sequence of subsets in order of their expected effectiveness (in the statistical sense) for the estimation of Φ.

In the following sections the main formulae estimating low-order invariants and seminvariants or relating phases to other phases and diffraction magnitudes are given.

2.2.5.3. Triplet relationships

The basic formula for the estimation of the triplet phase $\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}$ given the parameter $G = 2\sigma_3\sigma_2^{-3/2} \times R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{h}-\mathbf{k}}$ is Cochran's (1955) formula

$$P(\Phi) = [2\pi I_0(G)]^{-1} \exp(G\cos\Phi), \tag{2.2.5.6}$$

where $\sigma_n = \sum_{j=1}^{N} Z_j^n$, Z_j is the atomic number of the jth atom and I_n is the modified Bessel function of order n. In Fig. 2.2.5.1 the distribution $P(\Phi)$ is shown for different values of G.

The conditional probability distribution for $\varphi_{\mathbf{h}}$, given a set of $(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})$ and $G_j = 2\sigma_3\sigma_2^{-3/2}R_{\mathbf{h}}R_{\mathbf{k}_j}R_{\mathbf{h}-\mathbf{k}_j}$, is given (Karle & Hauptman, 1956; Karle & Karle, 1966) by

$$P(\varphi_{\mathbf{h}}) = [2\pi I_0(\alpha)]^{-1} \exp[\alpha\cos(\varphi_{\mathbf{h}} - \beta_{\mathbf{h}})], \tag{2.2.5.7}$$

where

$$\alpha^2 = \left[\sum_{j=1}^{r} G_{\mathbf{h},\mathbf{k}_j}\cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})\right]^2 + \left[\sum_{j=1}^{r} G_{\mathbf{h},\mathbf{k}_j}\sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})\right]^2, \tag{2.2.5.8}$$

$$\tan\beta_{\mathbf{h}} = \frac{\sum_j G_{\mathbf{h},\mathbf{k}_j}\sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j G_{\mathbf{h},\mathbf{k}_j}\cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}. \tag{2.2.5.9}$$

$\beta_{\mathbf{h}}$ is the most probable value for $\varphi_{\mathbf{h}}$. The variance of $\varphi_{\mathbf{h}}$ may be obtained from (2.2.5.7) and is given by

$$V_{\mathbf{h}} = \frac{\pi^2}{3} + [I_0(\alpha)]^{-1}\sum_{n=1}^{\infty}\frac{I_{2n}(\alpha)}{n^2} - 4[I_0(\alpha)]^{-1}\sum_{n=0}^{\infty}\frac{I_{2n+1}(\alpha)}{(2n+1)^2}, \tag{2.2.5.10}$$

which is plotted in Fig. 2.2.5.2.

Equation (2.2.5.9) is the so-called *tangent formula*. According to (2.2.5.10), the larger is α the more reliable is the relation $\varphi_{\mathbf{h}} = \beta_{\mathbf{h}}$.

For an equal-atom structure $\sigma_3\sigma_2^{-3/2} = N^{-1/2}$.

The basic conditional formula for sign determination of $E_{\mathbf{h}}$ in cs. crystals is Cochran & Woolfson's (1955) formula

$$P^+ = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\sum_{j=1}^{r}E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}\right), \tag{2.2.5.11}$$

where P^+ is the probability that $E_{\mathbf{h}}$ is positive and $\mathbf{k}$ ranges over the set of known values $E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$. The larger the absolute value of the argument of tanh, the more reliable is the phase indication.

An auxiliary formula exploiting all the $|E|$'s in reciprocal space in order to estimate a single Φ is the $B_{3,0}$ formula (Hauptman & Karle, 1958; Karle & Hauptman, 1958) given by

$$\begin{aligned}|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{-\mathbf{h}_1-\mathbf{h}_2}|\cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1+\mathbf{h}_2}) &\simeq C\langle(|E_{\mathbf{k}}|^p - \overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{k}}|^p - \overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^p - \overline{|E|^p})\rangle_{\mathbf{k}} \\ &\quad - \frac{2\sigma_6}{\sigma_4^{3/2}} + \frac{\sigma_8^{1/2}}{\sigma_4}(|E_{\mathbf{h}_1}|^2 + |E_{\mathbf{h}_2}|^2 + |E_{\mathbf{h}_1+\mathbf{h}_2}|^2)\ldots,\end{aligned} \tag{2.2.5.12}$$

where C is a constant which differs for cs. and ncs. crystals, $\overline{|E|^p}$ is the average value of $|E|^p$ and p is normally chosen to be some small number. Several modifications of (2.2.5.12) have been proposed (Hauptman, 1964, 1970; Karle, 1970*a*; Giacovazzo, 1977*b*).

A recent formula (Cascarano, Giacovazzo, Camalli *et al.*, 1984) exploits information contained within the second representation of Φ, that is to say, within the collection of special quintets (see Section 2.2.5.6):

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1+\mathbf{h}_2} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{k}},$$

where $\mathbf{k}$ is a free vector. The formula retains the same algebraic form as (2.2.5.6), but

$$G = \frac{2R_{\mathbf{h}_1}R_{\mathbf{h}_2}R_{\mathbf{h}_3}}{\sqrt{N}}(1 + Q), \tag{2.2.5.13}$$

where $[\mathbf{h}_3 = -(\mathbf{h}_1 + \mathbf{h}_2)]$,

$$Q = \sum_{\mathbf{k}} \frac{\sum_{i=1}^{'m} A_{\mathbf{k},i}/N}{1 + \left(\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{h}_3} + \sum_{i=1}^{'m} B_{\mathbf{k},i}\right)\Big/ 2N},$$

$$\begin{aligned}
A_{\mathbf{k},i} = \varepsilon_{\mathbf{k}}[&\varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i}(\varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i}) \\
&+ \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}(\varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i}) \\
&+ \varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}(\varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i})], \\
B_{\mathbf{k},i} = \varepsilon_{\mathbf{h}_1}[&\varepsilon_{\mathbf{k}}(\varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i}) \\
&+ \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}] \\
&+ \varepsilon_{\mathbf{h}_2}[\varepsilon_{\mathbf{k}}(\varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i}) \\
&+ \varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}] \\
&+ \varepsilon_{\mathbf{h}_3}[\varepsilon_{\mathbf{k}}(\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i}) \\
&+ \varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i}\varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}];
\end{aligned}$$

$\varepsilon = |E|^2 - 1$, $(\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{h}_3} + \sum_{i=1}^{'m} B_{\mathbf{k},i})$ is assumed to be zero if it is experimentally negative. The prime to the summation warns the reader that precautions have to be taken in order to avoid duplications in the contributions.

G may be positive or negative. In particular, if $G < 0$ the triplet is estimated negative.

The accuracy with which the value of Φ is estimated strongly depends on $\varepsilon_{\mathbf{k}}$. Thus, in practice, only a subset of reciprocal space (the reflections $\mathbf{k}$ with large values of ε) may be used for estimating Φ.

(2.2.5.13) proved to be quite useful in practice. Positive triplet cosines are ranked in order of reliability by (2.2.5.13) markedly better than by Cochran's parameters. Negative estimated triplet cosines may be excluded from the phasing process and may be used as a figure of merit for finding the correct solution in a multisolution procedure.

2.2.5.4. Triplet relationships using structural information

A strength of direct methods is that no knowledge of structure is required for their application. However, when some *a priori* information is available, it should certainly be a weakness of the methods not to make use of this knowledge. The conditional distribution of Φ given $R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{h}-\mathbf{k}}$ and the first three of the five kinds of *a priori* information described in Section 2.2.4.1 is (Main, 1976; Heinermann, 1977*a*)

$$P(\Phi) \simeq \frac{\exp[2QR_1R_2R_3\cos(\Phi - q)]}{2\pi I_0(2QR_1R_2R_3)}, \qquad (2.2.5.14)$$

where

$$Q\exp(iq) = \frac{\sum_{i=1}^{p} g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)}{\langle|F_{\mathbf{h}_1}|^2\rangle^{1/2}\langle|F_{\mathbf{h}_2}|^2\rangle^{1/2}\langle|F_{\mathbf{h}_3}|^2\rangle^{1/2}}.$$

$\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ stand for $\mathbf{h}$, $-\mathbf{k}$, $-\mathbf{h}+\mathbf{k}$, and R_1, R_2, R_3 for $R_{\mathbf{h}}$, $R_{\mathbf{k}}$, $R_{\mathbf{h}-\mathbf{k}}$. The quantities $\langle|F_{\mathbf{h}_i}|^2\rangle$ have been calculated in Section 2.2.4.1 according to different categories: $g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ is a suitable average of the product of three scattering factors for the ith atomic group, p is the number of atomic groups in the cell including those related by symmetry elements. We have the following categories.

(*a*) *No structural information*

(2.2.5.14) then reduces to (2.2.5.6).

(*b*) *Randomly positioned and randomly oriented atomic groups*

Then

$$g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j,k,l} f_j\, f_k\, f_l \langle\exp[2\pi i(\mathbf{h}_1\cdot\mathbf{r}_{kj} + \mathbf{h}_2\cdot\mathbf{r}_{lj})]\rangle_{\mathrm{R}},$$

where $\langle\ldots\rangle_{\mathrm{R}}$ means rotational average. The average of the exponential term extends over all orientations of the triangle formed by the atoms j, k and l, and is given (Hauptman, 1965) by

$$\begin{aligned}
B(z,t) &= \langle\exp[2\pi i(\mathbf{h}\cdot\mathbf{r} + \mathbf{h}'\cdot\mathbf{r}')]\rangle \\
&= \left(\frac{\pi}{2z}\right)^{1/2}\sum_{n=0}^{\infty}\frac{t^{2n}}{(n!)^2}J_{(4n+1)/2}(z),
\end{aligned}$$

where

$$z = 2\pi[q^2r^2 + 2qrq'r'\cos\varphi_q\cos\varphi_r + q'^2r'^2]^{1/2}$$

and

$$t = [2\pi^2 qrq'r'\sin\varphi_q\sin\varphi_r]/z;$$

q, q', r and r' are the magnitudes of $\mathbf{h}$, $\mathbf{h}'$, $\mathbf{r}$ and $\mathbf{r}'$, respectively; φ_q and φ_r are the angles $(\mathbf{h}, \mathbf{h}')$ and $(\mathbf{r}, \mathbf{r}')$, respectively.

(*c*) *Randomly positioned but correctly oriented atomic groups*

Then

$$\begin{aligned}
g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = &\sum_{s=1}^{m}\sum_{j,k,l} f_j\, f_k\, f_l \\
&\times \exp[2\pi i(\mathbf{h}_1\cdot\mathbf{R}_s\mathbf{r}_{kj} + \mathbf{h}_2\cdot\mathbf{R}_s\mathbf{r}_{lk})],
\end{aligned}$$

where the summations over j, k, l are taken over all the atoms in the ith group.

A modified expression for g_i has to be used in polar space groups for special triplets (Giacovazzo, 1988).

Translation functions [see Chapter 2.3; for an overview, see also Beurskens *et al.* (1987)] are also used to determine the position of a correctly oriented molecular fragment.

Such functions can work in direct space [expressed as Patterson convolutions (Buerger, 1959; Nordman, 1985) or electron-density convolutions (Rossmann *et al.*, 1964; Argos & Rossmann, 1980)] or in reciprocal space [expressed as correlation functions (Crowther & Blow, 1967; Karle, 1972; Langs, 1985) or residual functions (Rae, 1977)]. Both the probabilistic methods and the translation functions are quite efficient tools: the decision as to which one to use is often a personal choice.

(*d*) *Atomic groups correctly positioned*

Let p be the number of atoms with known position, q the number of atoms with unknown position, F_p and F_q the corresponding structure factors.

Tangent recycling methods (Karle, 1970*b*) may be used for recovering the complete crystal structure. The phase $\varphi_{p,\mathbf{h}}$ is accepted in the starting set as a useful approximation of $\varphi_{\mathbf{h}}$ if $|F_{p,\mathbf{h}}| > \eta|F_{\mathbf{h}}|$, where η is the fraction of the total scattering power contained in the fragment and where $|F_{\mathbf{h}}|$ is associated with $|E_{\mathbf{h}}| > 1.5$.

Tangent recycling methods are applied (Beurskens *et al.*, 1979) with greater effectiveness to difference s.f.'s $\Delta F = (|F| - |F_p|)\exp(i\varphi_p)$. The weighted tangent formula uses $\Delta F_{\mathbf{h}}$ values in order to convert them to more probable $F_{q,\mathbf{h}}$ values.

From a probabilistic point of view (Giacovazzo, 1983*a*; Camalli *et al.*, 1985) the distribution of $\varphi_{\mathbf{h}}$, given $E'_{p,\mathbf{h}}$ and some products $(E'_{\mathbf{k}} - E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}} - E'_{p,\mathbf{h}-\mathbf{k}})$, is the von Mises function

$$P(\varphi_{\mathbf{h}}|\ldots) = [2\pi I_0(\alpha)]^{-1}\exp[\alpha\cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})], \qquad (2.2.5.15)$$

where $\theta_{\mathbf{h}}$, the most probable value of $\varphi_{\mathbf{h}}$, is given by

$$\begin{aligned}
\tan\theta_{\mathbf{h}} &\simeq \alpha'_2/\alpha'_1, \qquad (2.2.5.16)\\
\alpha^2 &= \alpha'^2_1 + \alpha'^2_2
\end{aligned}$$

and

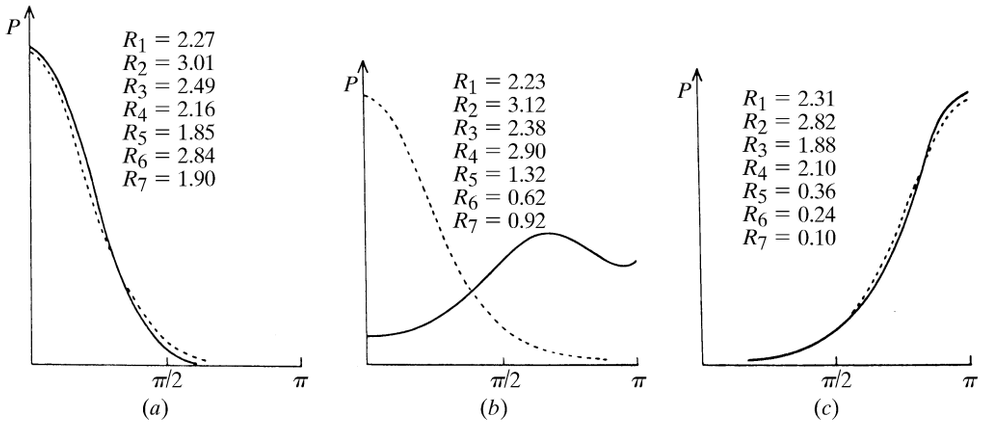


Fig. 2.2.5.3. Distributions (2.2.5.18) (solid curve) and (2.2.5.20) (dashed curve) for the indicated $|E|$ values in three typical cases.

$$\alpha'_1 = 2R'_{\mathbf{h}}\Big\{\mathscr{R}\Big[E'_{p,\,\mathbf{h}} + q^{-1/2}\textstyle\sum_{\mathbf{k}}(E'_{\mathbf{k}} - E'_{p,\,\mathbf{k}}) \times (E'_{\mathbf{h}-\mathbf{k}} - E'_{p,\,\mathbf{h}-\mathbf{k}})\Big]\Big\}$$

$$\alpha'_2 = 2R'_{\mathbf{h}}\Big\{\mathscr{I}\Big[E'_{p,\,\mathbf{h}} + q^{-1/2}\textstyle\sum_{\mathbf{k}}(E'_{\mathbf{k}} - E'_{p,\,\mathbf{k}}) \times (E'_{\mathbf{h}-\mathbf{k}} - E'_{p,\,\mathbf{h}-\mathbf{k}})\Big]\Big\}.$$

$\mathscr{R}$ and $\mathscr{I}$ stand for 'real and imaginary part of', respectively. Furthermore, $E' = F/\sum_q^{1/2}$ is a pseudo-normalized s.f. If no pair $(\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}})$ is known, then

$$\alpha'_1 = 2R'_{\mathbf{h}}R'_{p,\,\mathbf{h}}\cos\varphi_{p,\,\mathbf{h}}$$
$$\alpha'_2 = 2R'_{\mathbf{h}}R'_{p,\,\mathbf{h}}\sin\varphi_{p,\,\mathbf{h}}$$

and (2.2.5.15) reduces to Sim's (1959) equation

$$P(\varphi_{\mathbf{h}}) \simeq [2\pi I_0(G)]^{-1}\exp[G\cos(\varphi_{\mathbf{h}} - \varphi_{p,\,\mathbf{h}})], \qquad (2.2.5.17)$$

where $G = 2R'_{\mathbf{h}}R'_{p,\,\mathbf{h}}$. In this case $\varphi_{p,\,\mathbf{h}}$ is the most probable value of $\varphi_{\mathbf{h}}$.

(*e*) *Pseudotranslational symmetry is present*

Substructure and superstructure reflections are then described by different forms of the structure-factor equation (Böhme, 1982; Gramlich, 1984; Fan *et al.*, 1983), so that probabilistic formulae estimating triplet cosines derived on the assumption that atoms are uniformly dispersed in the unit cell cannot hold. In particular, the reliability of each triplet also depends on, besides $R_{\mathbf{h}}$, $R_{\mathbf{k}}$, $R_{\mathbf{h}-\mathbf{k}}$, the actual $\mathbf{h}$, $\mathbf{k}$, $\mathbf{h}-\mathbf{k}$ indices and on the nature of the pseudotranslation. It has been shown (Cascarano *et al.*, 1985*b*; Cascarano, Giacovazzo & Luić, 1987) that (2.2.5.7), (2.2.5.8), (2.2.5.9) still hold provided $G_{\mathbf{h},\,\mathbf{k}_j}$ is replaced by

$$G'_{\mathbf{h},\,\mathbf{k}_j} = \frac{2R_{\mathbf{h}}R_{\mathbf{k}_j}R_{\mathbf{h}-\mathbf{k}_j}}{\sqrt{N_{\mathbf{h},\,\mathbf{k}}}},$$

where factors E and n_i are defined according to Section 2.2.4.1,

$$N_{\mathbf{h},\mathbf{k}} = \frac{(\zeta_{\mathbf{h}}[\sigma_2]_p + [\sigma_2]_q)(\zeta_{\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)(\zeta_{\mathbf{h}-\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)}{\{(\beta/m)[\sigma_3]_p(n_1^2n_2^2n_3^2\ldots) + [\sigma_3]_q\}^2},$$

and β is the number of times for which

$$\mathbf{h}\mathbf{R}_s\cdot\mathbf{u}_1 \equiv 0 \pmod 1 \qquad \mathbf{h}\mathbf{R}_s\cdot\mathbf{u}_2 \equiv 0 \pmod 1 \qquad \mathbf{h}\mathbf{R}_s\cdot\mathbf{u}_3 \equiv 0 \pmod 1\ldots$$
$$\mathbf{k}\mathbf{R}_s\cdot\mathbf{u}_1 \equiv 0 \pmod 1 \qquad \mathbf{k}\mathbf{R}_s\cdot\mathbf{u}_2 \equiv 0 \pmod 1 \qquad \mathbf{k}\mathbf{R}_s\cdot\mathbf{u}_3 \equiv 0 \pmod 1\ldots$$
$$(\mathbf{h}-\mathbf{k})\mathbf{R}_s\cdot\mathbf{u}_1 \equiv 0 \pmod 1 \qquad (\mathbf{h}-\mathbf{k})\mathbf{R}_s\cdot\mathbf{u}_2 \equiv 0 \pmod 1$$
$$(\mathbf{h}-\mathbf{k})\mathbf{R}_s\cdot\mathbf{u}_3 \equiv 0 \pmod 1\ldots$$

are simultaneously satisfied when s varies from 1 to m. The above formulae have been generalized (Cascarano *et al.*, 1988*b*) to the case in which deviations both of replacive and of displacive type from ideal pseudo-translational symmetry occur.

2.2.5.5. Quartet phase relationships

In early papers (Hauptman & Karle, 1953; Simerska, 1956) the phase

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} - \varphi_{\mathbf{h}+\mathbf{k}+\mathbf{l}}$$

was always expected to be zero. Schenk (1973*a*,*b*) [see also Hauptman (1974)] suggested that Φ primarily depends on the seven magnitudes: $R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{l}}, R_{\mathbf{h}+\mathbf{k}+\mathbf{l}}$, called *basis magnitudes*, and $R_{\mathbf{h}+\mathbf{k}}, R_{\mathbf{h}+\mathbf{l}}, R_{\mathbf{k}+\mathbf{l}}$, called *cross magnitudes*.

The conditional probability of Φ in $P1$ given seven magnitudes $(R_1 = R_{\mathbf{h}}, \ldots, R_4 = R_{\mathbf{h}+\mathbf{k}+\mathbf{l}}, R_5 = R_{\mathbf{h}+\mathbf{k}}, R_6 = R_{\mathbf{h}+\mathbf{l}}, R_7 = R_{\mathbf{k}+\mathbf{l}})$ according to Hauptman (1975) is

$$P_7(\Phi) = \frac{1}{L}\exp(-2B\cos\Phi)I_0(2\sigma_3\sigma_2^{-3/2}R_5Y_5) \times I_0(2\sigma_3\sigma_2^{-3/2}R_6Y_6)I_0(2\sigma_3\sigma_2^{-3/2}R_7Y_7),$$

where L is a suitable normalizing constant which can be derived numerically,

$$B = \sigma_2^{-3}(3\sigma_3^2 - \sigma_2\sigma_4)R_1R_2R_3R_4$$
$$Y_5 = [R_1^2R_2^2 + R_3^2R_4^2 + 2R_1R_2R_3R_4\cos\Phi]^{1/2}$$
$$Y_6 = [R_3^2R_1^2 + R_2^2R_4^2 + 2R_1R_2R_3R_4\cos\Phi]^{1/2}$$
$$Y_7 = [R_2^2R_3^2 + R_1^2R_4^2 + 2R_1R_2R_3R_4\cos\Phi]^{1/2}.$$

For equal atoms $\sigma_2^{-3}(3\sigma_3^2 - \sigma_2\sigma_4) = 2/N$. Denoting

$$C = R_1R_2R_3R_4/N,$$
$$Z_5 = 2Y_5/\sqrt{N}, \quad Z_6 = 2Y_6/\sqrt{N}, \quad Z_7 = 2Y_7/\sqrt{N}$$

gives

$$P_7(\Phi) = \frac{1}{L}\exp(-4C\cos\Phi) \times I_0(R_5Z_5)I_0(R_6Z_6)I_0(R_7Z_7). \qquad (2.2.5.18)$$

Fig. 2.2.5.3 shows the distribution (2.2.5.18) for three typical cases. It is clear from the figure that the cosine estimated near π or in the middle range will be in poorer agreement with the true values than the cosine near 0 because of the relatively larger values of the variance. In principle, however, the formula is able to estimate negative or enantiomorph-sensitive quartet cosines from the seven magnitudes.

In the cs. case (2.2.5.18) is replaced (Hauptman & Green, 1976) by

Table 2.2.5.1. *List of quartets symmetry equivalent to* $\Phi = \Phi_1$ *in the class mmm*

Quartets	Basis vectors				Cross vectors		
Φ_1	$(1, 2, 3)$	$(\bar{1}, 5, \bar{3})$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(0, \bar{3}, 11)$	$(\bar{2}, 0, 5)$
Φ_2	$(\bar{1}, 2, 3)$	$(1, 5, \bar{3})$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(\bar{2}, \bar{3}, 11)$	$(0, 0, 5)$
Φ_3	$(1, 2, \bar{3})$	$(\bar{1}, 5, 3)$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(0, \bar{3}, 5)$	$(\bar{2}, 0, 11)$
Φ_4	$(\bar{1}, 2, \bar{3})$	$(1, 5, 3)$	$(\bar{1}, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(0, 7, 0)$	$(\bar{2}, \bar{3}, 5)$	$(0, 0, 11)$
Φ_5	$(\bar{1}, 2, 3)$	$(\bar{1}, 5, \bar{3})$	$(1, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(\bar{2}, 7, 0)$	$(0, \bar{3}, 11)$	$(0, 0, 5)$
Φ_6	$(1, 2, 3)$	$(\bar{1}, \bar{5}, \bar{3})$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(0, 7, 11)$	$(\bar{2}, 0, 5)$
Φ_7	$(\bar{1}, 2, 3)$	$(1, \bar{5}, \bar{3})$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(\bar{2}, 7, 11)$	$(0, 0, 5)$
Φ_8	$(\bar{1}, 2, \bar{3})$	$(\bar{1}, 5, 3)$	$(1, \bar{5}, 8)$	$(1, \bar{2}, \bar{8})$	$(\bar{2}, 7, 0)$	$(0, \bar{3}, 5)$	$(0, 0, 11)$
Φ_9	$(1, 2, \bar{3})$	$(\bar{1}, \bar{5}, 3)$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(0, 7, 5)$	$(\bar{2}, 0, 11)$
Φ_{10}	$(\bar{1}, 2, \bar{3})$	$(1, \bar{5}, 3)$	$(\bar{1}, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(0, \bar{3}, 0)$	$(\bar{2}, 7, 5)$	$(0, 0, 11)$
Φ_{11}	$(\bar{1}, 2, 3)$	$(\bar{1}, \bar{5}, \bar{3})$	$(1, 5, 8)$	$(1, \bar{2}, \bar{8})$	$(\bar{2}, \bar{3}, 0)$	$(0, 7, 11)$	$(0, 0, 5)$

$$P^{\pm} \simeq \frac{1}{L}\exp(\mp 2C)\cosh(R_5 Z_5^{\pm}) \times \cosh(R_6 Z_6^{\pm})\cosh(R_7 Z_7^{\pm}), \qquad (2.2.5.19)$$

where $P^{\pm}$ is the probability that the sign of $E_1E_2E_3E_4$ is positive or negative, and

$$Z_5^{\pm} = \frac{1}{N^{1/2}}(R_1R_2 \pm R_3R_4),$$
$$Z_6^{\pm} = \frac{1}{N^{1/2}}(R_1R_3 \pm R_2R_4),$$
$$Z_7^{\pm} = \frac{1}{N^{1/2}}(R_1R_4 \pm R_2R_3).$$

The normalized probability may be derived by $P^+/(P^+ + P^-)$. More simple probabilistic formulae were derived independently by Giacovazzo (1975, 1976):

$$P_7(\Phi) = [2\pi I_0(G)]^{-1}\exp(G\cos\Phi), \qquad (2.2.5.20)$$

where

$$G = \frac{2C(1 + \varepsilon_5 + \varepsilon_6 + \varepsilon_7)}{1 + Q/(2N)} \qquad (2.2.5.21)$$

$$Q = (\varepsilon_1\varepsilon_2 + \varepsilon_3\varepsilon_4)\varepsilon_5 + (\varepsilon_1\varepsilon_3 + \varepsilon_2\varepsilon_4)\varepsilon_6 + (\varepsilon_1\varepsilon_4 + \varepsilon_2\varepsilon_3)\varepsilon_7 \qquad (2.2.5.22)$$

and $\varepsilon_i = (|E_i|^2 - 1)$. Q is never allowed to be negative.

According to (2.2.5.20) $\cos\Phi$ is expected to be positive or negative according to whether $(\varepsilon_5 + \varepsilon_6 + \varepsilon_7 + 1)$ is positive or negative: the larger is C, the more reliable is the phase indication. For $N \geq 150$, (2.2.5.18) and (2.2.5.20) are practically equivalent in all cases. If N is small, (2.2.5.20) is in good agreement with (2.2.5.18) for quartets strongly defined as positive or negative, but in poor agreement for enantiomorph-sensitive quartets (see Fig. 2.2.5.3).

In cs. cases the sign probability for $E_1E_2E_3E_4$ is

$$P^+ = \tfrac{1}{2} + \tfrac{1}{2}\tanh(G/2), \qquad (2.2.5.23)$$

where G is defined by (2.2.5.21).

All three cross magnitudes are not always in the set of measured reflections. From marginal distributions the following formulae arise (Giacovazzo, 1977*c*; Heinermann, 1977*b*):

(*a*) in the ncs. case, if R_7, or R_6 and R_7, or R_5 and R_6 and R_7, are not in the measurements, then (2.2.5.18) is replaced by

$$P(\Phi|R_1, \ldots, R_6) \simeq \frac{1}{L'}\exp(-2C\cos\Phi)I_0(R_5Z_5)I_0(R_6Z_6),$$

or

$$P(\Phi|R_1, \ldots, R_5) \simeq \frac{1}{L''}I_0(R_5Z_5),$$

or

$$P(\Phi|R_1, \ldots, R_4) \simeq \frac{1}{L'''}\exp(2C\cos\Phi),$$

respectively.

(*b*) in the same situations, we have for cs. cases

$$P^{\pm} \simeq \frac{1}{L'}\exp(\mp C)\cosh(R_5Z_5^{\pm})\cosh(R_6Z_6^{\pm}),$$

or

$$P^{\pm} \simeq \frac{1}{L''}\cosh(R_5Z_5^{\pm}),$$

or

$$P^{\pm} = \frac{1}{L'''}\exp(\pm C) \simeq 0.5 + 0.5\tanh(\pm C),$$

respectively.

Equations (2.2.5.20) and (2.2.5.23) are easily modifiable when some cross magnitudes are not in the measurements. If R_i is not measured then (2.2.5.20) or (2.2.5.23) are still valid provided that in G it is assumed that $\varepsilon_i = 0$. For example, if R_7 and R_6 are not in the data then (2.2.5.21) and (2.2.5.22) become

$$G = \frac{2C(1 + \varepsilon_5)}{1 + Q/(2N)}, \qquad Q = (\varepsilon_1\varepsilon_2 + \varepsilon_3\varepsilon_4)\varepsilon_5.$$

In space groups with symmetry higher than $P\bar{1}$ more symmetry-equivalent quartets can exist of the type

$$\psi = \varphi_{\mathbf{h}\boldsymbol{R}_\alpha} + \varphi_{\mathbf{k}\boldsymbol{R}_\beta} + \varphi_{\mathbf{l}\boldsymbol{R}_\gamma} + \varphi_{(\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}})\boldsymbol{R}_\delta},$$

where $\boldsymbol{R}_\alpha, \boldsymbol{R}_\beta, \boldsymbol{R}_\gamma, \boldsymbol{R}_\delta$ are rotation matrices of the space group. The set $\{\psi\}$ is called the *first representation* of Φ. In this case Φ primarily depends on more than seven magnitudes. For example, let us consider in *Pmmm* the quartet

$$\Phi = \varphi_{123} + \varphi_{\bar{1}5\bar{3}} + \varphi_{\bar{1}\bar{5}8} + \varphi_{1\bar{2}\bar{8}}.$$

Quartets symmetry equivalent to Φ and respective cross terms are given in Table 2.2.5.1.

Experimental tests on the application of the representation concept to quartets have been made (Busetta *et al.*, 1980). It was shown that quartets with more than three cross magnitudes are more accurately estimated than other quartets. Also, quartets with a cross reflection which is systematically absent were shown to be of significant importance in direct methods. In this context it is noted that systematically absent reflections are not usually included in the set of diffraction data. This custom, not exceptionable when only triplet relations are used, can give rise to a loss of information when quartets are used. In fact the usual programs of direct methods discard quartets as soon as one of the cross reflections is not measured, so that systematic absences are

dealt with in the same manner as those reflections which are outside the sphere of measurements.

2.2.5.6. Quintet phase relationships

A quintet phase

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} + \varphi_{\overline{\mathbf{h+k+l+m}}}$$

may be considered as the sum of three suitable triplets or the sum of a triplet and a quartet, *i.e.*

$$\Phi = (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h+k}}) + (\varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} - \varphi_{\mathbf{l+m}}) + (\varphi_{\mathbf{h+k}} + \varphi_{\mathbf{l+m}} + \varphi_{\overline{\mathbf{h+k+l+m}}})$$

or

$$\Phi = (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h+k}}) + (\varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} + \varphi_{\overline{\mathbf{h+k+l+m}}} + \varphi_{\mathbf{h+k}}).$$

It depends primarily on 15 magnitudes: the five *basis magnitudes*

$$R_{\mathbf{h}}, \quad R_{\mathbf{k}}, \quad R_{\mathbf{l}}, \quad R_{\mathbf{m}}, \quad R_{\mathbf{h+k+l+m}},$$

and the ten *cross magnitudes*

$$R_{\mathbf{h+k}}, \quad R_{\mathbf{h+l}}, \quad R_{\mathbf{h+m}}, \quad R_{\mathbf{k+l+m}}, \quad R_{\mathbf{k+l}},$$
$$R_{\mathbf{k+m}}, \quad R_{\mathbf{h+l+m}}, \quad R_{\mathbf{l+m}}, \quad R_{\mathbf{h+k+m}}, \quad R_{\mathbf{h+k+l}}.$$

In the following we will denote

$$R_1 = R_{\mathbf{h}}, \quad R_2 = R_{\mathbf{k}}, \ldots, \quad R_{15} = R_{\mathbf{h+k+l}}.$$

Conditional distributions of Φ in $P1$ and $P\bar{1}$ given the 15 magnitudes have been derived by several authors and allow in favourable circumstances in ncs. space groups the quintets having Φ near 0 or near π or near $\pm\pi/2$ to be identified. Among others, we remember:

(*a*) the semi-empirical expression for $P_{15}(\Phi)$ suggested by Van der Putten & Schenk (1977):

$$P(\Phi|\ldots) \simeq \frac{1}{L}\exp\left[\left(6 - \sum_{j=6}^{15} R_j^2\right) 2C\cos\Phi\right] \prod_{j=6}^{15} I_0(2R_jY_j),$$

where

$$C = N^{-3/2}R_1R_2R_3R_4R_5$$

and Y_j is an expression related to the jth of the ten quartets connected with the quintet Φ;

(*b*) the formula by Fortier & Hauptman (1977), valid in $P\bar{1}$, which is able to predict the sign of a quintet by means of an expression which involves a summation over 1024 sets of signs;

(*c*) the expression by Giacovazzo (1977*d*), according to which

$$P_{15}(\Phi) \simeq [2\pi I_0(G)]^{-1} \exp(G\cos\Phi), \qquad (2.2.5.24)$$

where

$$G = \frac{2C}{1 + 6(N)^{1/2}}\left[\frac{1 + A + B}{1 + D/(2N)}\right] \qquad (2.2.5.25)$$

and where

$$A = \sum_{i=6}^{15} \varepsilon_i,$$

$$B = \varepsilon_6\varepsilon_{13} + \varepsilon_6\varepsilon_{15} + \varepsilon_6\varepsilon_{14} + \varepsilon_7\varepsilon_{11} + \varepsilon_7\varepsilon_{15} + \varepsilon_7\varepsilon_{12} + \varepsilon_8\varepsilon_{10} + \varepsilon_8\varepsilon_{14} + \varepsilon_8\varepsilon_{12} + \varepsilon_{10}\varepsilon_{15} + \varepsilon_{10}\varepsilon_9 + \varepsilon_{11}\varepsilon_{14} + \varepsilon_{11}\varepsilon_9 + \varepsilon_{13}\varepsilon_9 + \varepsilon_{13}\varepsilon_{12},$$

$$D = \varepsilon_1\varepsilon_2\varepsilon_6 + \varepsilon_1\varepsilon_3\varepsilon_7 + \varepsilon_1\varepsilon_4\varepsilon_8 + \varepsilon_1\varepsilon_5\varepsilon_9 + \varepsilon_1\varepsilon_{10}\varepsilon_{15} + \varepsilon_1\varepsilon_{11}\varepsilon_{14} + \varepsilon_1\varepsilon_{13}\varepsilon_{12} + \varepsilon_2\varepsilon_3\varepsilon_{10} + \varepsilon_2\varepsilon_4\varepsilon_{11} + \varepsilon_2\varepsilon_5\varepsilon_{12} + \varepsilon_2\varepsilon_7\varepsilon_{15} + \varepsilon_2\varepsilon_8\varepsilon_{14} + \varepsilon_2\varepsilon_{13}\varepsilon_9 + \varepsilon_3\varepsilon_4\varepsilon_{13} + \varepsilon_3\varepsilon_5\varepsilon_{14} + \varepsilon_3\varepsilon_6\varepsilon_{15} + \varepsilon_3\varepsilon_8\varepsilon_{12} + \varepsilon_3\varepsilon_{11}\varepsilon_9 + \varepsilon_4\varepsilon_5\varepsilon_{15} + \varepsilon_4\varepsilon_6\varepsilon_{14} + \varepsilon_4\varepsilon_7\varepsilon_{12} + \varepsilon_4\varepsilon_{10}\varepsilon_9 + \varepsilon_5\varepsilon_6\varepsilon_{13} + \varepsilon_5\varepsilon_7\varepsilon_{11} + \varepsilon_5\varepsilon_8\varepsilon_{10}.$$

For cs. cases (2.2.5.24) reduces to

$$P^+ \simeq 0.5 + 0.5\tanh(G/2). \qquad (2.2.5.26)$$

Positive or negative quintets may be identified according to whether G is larger or smaller than zero.

If R_i is not measured then (2.2.5.24) and (2.2.5.25) are still valid provided that in (2.2.5.25) $\varepsilon_i = 0$.

If the symmetry is higher than in $P\bar{1}$ then more symmetry-equivalent quintets can exist of the type

$$\psi = \varphi_{\mathbf{h}\boldsymbol{R}_\alpha} + \varphi_{\mathbf{k}\boldsymbol{R}_\beta} + \varphi_{\mathbf{l}\boldsymbol{R}_\gamma} + \varphi_{\mathbf{m}\boldsymbol{R}_\delta} + \varphi_{(\overline{\mathbf{h+k+l+m}})\boldsymbol{R}_\varepsilon},$$

where $\boldsymbol{R}_\alpha, \ldots, \boldsymbol{R}_\varepsilon$ are rotation matrices of the space groups. The set $\{\psi\}$ is called the first representation of Φ. In this case Φ primarily depends on more than 15 magnitudes which all have to be taken into account for a careful estimation of Φ (Giacovazzo, 1980*a*).

A wide use of quintet invariants in direct-methods procedures is prevented for two reasons: (*a*) the large correlation of positive quintet cosines with positive triplets; (*b*) the large computing time necessary for their estimation [quintets are phase relationships of order $1/(N\sqrt{N})$, so a large number of quintets have to be estimated in order to pick up a sufficient percentage of reliable ones].

2.2.5.7. Determinantal formulae

In a crystal structure with N identical atoms the joint probability distribution of n normalized s.f.'s $E_{\mathbf{h}_1+\mathbf{k}}, E_{\mathbf{h}_2+\mathbf{k}}, \ldots, E_{\mathbf{h}_n+\mathbf{k}}$ under the following conditions:

(*a*) the structure is kept fixed whereas $\mathbf{k}$ is the primitive random variable;

(*b*) $E_{\mathbf{h}_i-\mathbf{h}_j}$, $i, j = 1, \ldots, n$, have values which are known *a priori*;

is given (Tsoucaris, 1970) [see also Castellano *et al.* (1973) and Heinermann *et al.* (1979)] by

$$P(E_1, E_2, \ldots, E_n) = (2\pi)^{-n/2} D_n^{-1/2} \exp(-\tfrac{1}{2}Q_n) \qquad (2.2.5.27)$$

for cs. structures and

$$P(E_1, E_2, \ldots, E_n) = (2\pi)^{-n} D_n^{-1/2} \exp(-Q_n) \qquad (2.2.5.28)$$

for ncs. structures. In (2.2.5.27) and (2.2.5.28) we have denoted

$$D_n = \lambda, \qquad Q_n = \sum_{p,q=1}^{n} \Lambda_{pq} E_p E_q^*$$
$$E_j = E_{\mathbf{h}_j+\mathbf{k}}, \qquad U_{pq} = U_{\mathbf{h}_p-\mathbf{h}_q}, \qquad j, p, q = 1, \ldots, n.$$

Λ_{pq} is an element of $\boldsymbol{\lambda}^{-1}$, and $\boldsymbol{\lambda}$ is the covariance matrix with elements

$$\langle E_{\mathbf{h}_p+\mathbf{k}} E_{\mathbf{h}_q+\mathbf{k}} \rangle = U_{\mathbf{h}_p-\mathbf{h}_q}$$

$$\boldsymbol{\lambda} = \begin{vmatrix} 1 & U_{12} & \dots & U_{1q} & \dots & U_{1n} \\ U_{21} & 1 & \dots & U_{2q} & \dots & U_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{p1} & U_{p2} & \dots & U_{pq} & \dots & U_{pn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{n1} & U_{n2} & \dots & U_{nq} & \dots & 1 \end{vmatrix}.$$

λ is a K–H determinant: therefore $D_n \geq 0$. Let us call

$$\Delta_{n+1} = \frac{1}{N} \begin{vmatrix} 1 & U_{12} & \dots & U_{1n} & E_{\mathbf{h}_1+\mathbf{k}} \\ U_{21} & 1 & \dots & U_{2n} & E_{\mathbf{h}_2+\mathbf{k}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ U_{n1} & U_{n2} & \dots & 1 & E_{\mathbf{h}_n+\mathbf{k}} \\ E_{-\mathbf{h}_1-\mathbf{k}} & E_{-\mathbf{h}_2-\mathbf{k}} & \dots & E_{-\mathbf{h}_n-\mathbf{k}} & N \end{vmatrix};$$

the K–H determinant obtained by adding to $\boldsymbol{\lambda}$ the last column and line formed by $E_1, E_2, \dots, E_n$, and $E_1^*, E_2^*, \dots, E_n^*$, respectively. Then (2.2.5.27) and (2.2.5.28) may be written

$$P(E_1, E_2, \dots, E_n) = (2\pi)^{-n/2} D_n^{-1/2} \exp\left[N \frac{\Delta_{n+1} - D_n}{2D_n}\right] \quad (2.2.5.29)$$

and

$$P(E_1, E_2, \dots, E_n) = (2\pi)^{-n} D_n^{-1/2} \exp\left[N \frac{\Delta_{n+1} - D_n}{D_n}\right], \quad (2.2.5.30)$$

respectively. Because D_n is a constant, the maximum values of the conditional joint probabilities (2.2.5.29) and (2.2.5.30) are obtained when Δ_{n+1} is a maximum. Thus the *maximum determinant rule* may be stated (Tsoucaris, 1970; Lajzérowicz & Lajzérowicz, 1966): among all sets of phases which are compatible with the inequality

$$\Delta_{n+1}(E_1, E_2, \dots, E_n) \geq 0$$

the most probable one is that which leads to a maximum value of Δ_{n+1}.

If only one phase, *i.e.* φ_q, is unknown whereas all other phases and moduli are known then (de Rango *et al.*, 1974; Podjarny *et al.*, 1976) for cs. crystals

$$P^{\pm}(E_q) \simeq 0.5 + 0.5 \tanh \left\{ \pm |E_q| \sum_{\substack{p=1 \\ p \neq q}}^{n} \Lambda_{pq} E_p \right\}, \quad (2.2.5.31)$$

and for ncs. crystals

$$P(\varphi_q) = [2\pi I_0(G_q)]^{-1} \exp\{G_q \cos(\varphi_q - \theta_q)\}, \quad (2.2.5.32)$$

where

$$G_q \exp(i\theta_q) = 2|E_q| \sum_{p \neq q=1}^{n} \Lambda_{pq} E_p.$$

Equations (2.2.5.31) and (2.2.5.32) generalize (2.2.5.11) and (2.2.5.7), respectively, and reduce to them for $n = 3$. Fourth-order determinantal formulae estimating triplet invariants in cs. and ncs. crystals, and making use of the entire data set, have recently been secured (Karle, 1979, 1980*a*).

Advantages, limitations and applications of determinantal formulae can be found in the literature (Heinermann *et al.*, 1979; de Rango *et al.*, 1975, 1985). Taylor *et al.* (1978) combined K–H determinants with a magic-integer approach. The computing time, however, was larger than that required by standard computing techniques. The use of K–H matrices has been made faster and more effective by de Gelder *et al.* (1990) (see also de Gelder, 1992). They developed a phasing procedure (*CRUNCH*) which uses random phases as starting points for the maximization of the K–H determinants.

2.2.5.8. Algebraic relationships for structure seminvariants

According to the representations method (Giacovazzo, 1977*a*, 1980*a*,*b*):

(i) any s.s. Φ may be estimated *via* one or more s.i.'s $\{\psi\}$, whose values differ from Φ by a constant arising because of symmetry;

(ii) two types of s.s.'s exist, first-rank and second-rank s.s.'s, with different algebraic properties:

(iii) conditions characterizing s.s.'s of first rank for any space group may be expressed in terms of seminvariant moduli and seminvariantly associated vectors. For example, for all the space groups with point group 422 [Hauptman–Karle group $(h + k, l)$ $P(2, 2)$] the one-phase s.s.'s of first rank are characterized by

$$(h, k, l) \equiv 0 \bmod (2, 2, 0) \text{ or } (2, 0, 2) \text{ or } (0, 2, 2)$$
$$(h \pm k, l) \equiv 0 \bmod (0, 2) \text{ or } (2, 0).$$

The more general expressions for the s.s.'s of first rank are

(*a*) $\Phi = \varphi_{\mathbf{u}} = \varphi_{\mathbf{h}(\boldsymbol{I}-\boldsymbol{R}_\alpha)}$ for one-phase s.s.'s;

(*b*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} = \varphi_{\mathbf{h}_1-\mathbf{h}_2\boldsymbol{R}_\beta} + \varphi_{\mathbf{h}_2-\mathbf{h}_1\boldsymbol{R}_\alpha}$ for two-phase s.s.'s;

(*c*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} + \varphi_{\mathbf{u}_3} = \varphi_{\mathbf{h}_1-\mathbf{h}_2\boldsymbol{R}_\beta} + \varphi_{\mathbf{h}_2-\mathbf{h}_3\boldsymbol{R}_\gamma} + \varphi_{\mathbf{h}_3-\mathbf{h}_1\boldsymbol{R}_\alpha}$ for three-phase s.s.'s;

(*d*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} + \varphi_{\mathbf{u}_3} + \varphi_{\mathbf{u}_4} = \varphi_{\mathbf{h}_1-\mathbf{h}_2\boldsymbol{R}_\beta} + \varphi_{\mathbf{h}_2-\mathbf{h}_3\boldsymbol{R}_\gamma} + \varphi_{\mathbf{h}_3-\mathbf{h}_4\boldsymbol{R}_\delta} + \varphi_{\mathbf{h}_4-\mathbf{h}_1\boldsymbol{R}_\alpha}$

for four-phase s.s.'s; *etc.*

In other words:

(*a*) $\varphi_{\mathbf{u}}$ is an s.s. of first rank if at least one $\mathbf{h}$ and at least one rotation matrix $\boldsymbol{R}_\alpha$ exist such that $\mathbf{u} = \mathbf{h}(\boldsymbol{I} - \boldsymbol{R}_\alpha)$. $\varphi_{\mathbf{u}}$ may be estimated *via* the special triplet invariants

$$\{\psi\} = \varphi_{\mathbf{u}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\boldsymbol{R}_\alpha}. \quad (2.2.5.33)$$

The set $\{\psi\}$ is called the *first representation* of $\varphi_{\mathbf{u}}$.

(*b*) $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$ is an s.s. of first rank if at least two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ and two rotation matrices $\boldsymbol{R}_\alpha$ and $\boldsymbol{R}_\beta$ exist such that

$$\begin{cases} \mathbf{u}_1 = \mathbf{h}_1 - \mathbf{h}_2\boldsymbol{R}_\beta \\ \mathbf{u}_2 = \mathbf{h}_2 - \mathbf{h}_1\boldsymbol{R}_\alpha. \end{cases} \quad (2.2.5.34)$$

Φ may then be estimated *via* the special quartet invariants

$$\{\psi\} = \varphi_{\mathbf{u}_1\boldsymbol{R}_\alpha} + \varphi_{\mathbf{u}_2} - \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_2\boldsymbol{R}_\beta\boldsymbol{R}_\alpha} \quad (2.2.5.35a)$$

and

$$\{\psi\} = \{\varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2\boldsymbol{R}_\beta} - \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_1\boldsymbol{R}_\alpha\boldsymbol{R}_\beta}\}. \quad (2.2.5.35b)$$

For example, $\Phi = \varphi_{123} + \varphi_{\bar{7}\bar{2}\bar{5}}$ in $P2_1$ may be estimated *via*

$$\{\psi\} = \varphi_{123} + \varphi_{\bar{7}\bar{2}\bar{5}} - \varphi_{\bar{3}K\bar{1}} + \varphi_{3K1}$$

and

$$\{\psi\} = \varphi_{123} + \varphi_{7\bar{2}5} - \varphi_{4K4} + \varphi_{\bar{4}K\bar{4}},$$

where K is a free index.

The set of special quartets (2.2.5.35*a*) and (2.2.5.35*b*) constitutes the *first representations* of Φ.

Structure seminvariants of the second rank can be characterized as follows: suppose that, for a given seminvariant Φ, it is not possible to find a vectorial index $\mathbf{h}$ and a rotation matrix $\boldsymbol{R}_\alpha$ such that $\Phi - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\boldsymbol{R}_\alpha}$ is a structure invariant. Then Φ is a structure

seminvariant of the second rank and a set of structure invariants ψ can certainly be formed, of type

$$\{\psi\} = \Phi + \varphi_{\mathbf{h}\boldsymbol{R}_p} - \varphi_{\mathbf{h}\boldsymbol{R}_q} + \varphi_{\mathbf{l}\boldsymbol{R}_i} - \varphi_{\mathbf{l}\boldsymbol{R}_j},$$

by means of suitable indices $\mathbf{h}$ and $\mathbf{l}$ and rotation matrices $\boldsymbol{R}_p$, $\boldsymbol{R}_q$, $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$. As an example, for symmetry class 222, φ_{240} or φ_{024} or φ_{204} are s.s.'s of the first rank while φ_{246} is an s.s. of the second rank.

The procedure may easily be generalized to s.s.'s of any order of the first and of the second rank. So far only the role of one-phase and two-phase s.s.'s of the first rank in direct procedures is well documented (see references quoted in Sections 2.2.5.9 and 2.2.5.10).

2.2.5.9. Formulae estimating one-phase structure seminvariants of the first rank

Let $E_{\mathbf{H}}$ be our one-phase s.s. of the first rank, where

$$\mathbf{H} = \mathbf{h}(\boldsymbol{I} - \boldsymbol{R}_n). \tag{2.2.5.36}$$

In general, more than one rotation matrix $\boldsymbol{R}_n$ and more than one vector $\mathbf{h}$ are compatible with (2.2.5.36). The set of special triplets

$$\{\psi\} = \{\varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\boldsymbol{R}_n}\}$$

is the first representation of $E_{\mathbf{H}}$. In cs. space groups the probability that $E_{\mathbf{H}} > 0$, given $|E_{\mathbf{H}}|$ and the set $\{|E_{\mathbf{h}}|\}$, may be estimated (Hauptman & Karle, 1953; Naya *et al.*, 1964; Cochran & Woolfson, 1955) by

$$P^+(E_{\mathbf{H}}) \simeq 0.5 + 0.5 \tanh \sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n}(-1)^{2\mathbf{h}\cdot\mathbf{T}_n}, \tag{2.2.5.37}$$

where

$$G_{\mathbf{h},\,n} = |E_{\mathbf{H}}|\varepsilon_{\mathbf{h}}/(2\sqrt{N}), \text{ and } \varepsilon = |E|^2 - 1.$$

In (2.2.5.37), the summation over n goes within the set of matrices $\boldsymbol{R}_n$ for which (2.2.5.35*a*,*b*) is compatible, and $\mathbf{h}$ varies within the set of vectors which satisfy (2.2.5.36) for each $\boldsymbol{R}_n$. Equation (2.2.5.36) is actually a generalized way of writing the so-called $\sum_1$ relationships (Hauptman & Karle, 1953).

If $\varphi_{\mathbf{H}}$ is a phase restricted by symmetry to $\theta_{\mathbf{H}}$ and $\theta_{\mathbf{H}} + \pi$ in an ncs. space group then (Giacovazzo, 1978)

$$P(\varphi_{\mathbf{H}} = \theta_{\mathbf{H}}) \simeq 0.5 + 0.5 \tanh \left\{ \sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n} \cos(\theta_{\mathbf{H}} - 2\pi\mathbf{h}\cdot\mathbf{T}_n) \right\}. \tag{2.2.5.38}$$

If $\varphi_{\mathbf{H}}$ is a general phase then $\varphi_{\mathbf{H}}$ is distributed according to

$$P(\varphi_{\mathbf{H}}) \simeq \frac{1}{L} \exp\{\alpha \cos(\varphi_{\mathbf{H}} - \theta_{\mathbf{H}})\},$$

where

$$\tan\theta_{\mathbf{H}} = \frac{\left(\sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n} \sin 2\pi\mathbf{h}\cdot\mathbf{T}_n\right)}{\left(\sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n} \cos 2\pi\mathbf{h}\cdot\mathbf{T}_n\right)} \tag{2.2.5.39}$$

with a reliability measured by

$$\alpha = \left\{ \left(\sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n} \sin 2\pi\mathbf{h}\cdot\mathbf{T}_n \right)^2 + \left(\sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n} \cos 2\pi\mathbf{h}\cdot\mathbf{T}_n \right)^2 \right\}^{1/2}.$$

The second representation of $\varphi_{\mathbf{H}}$ is the set of special quintets

$$\{\psi\} = \{\varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\boldsymbol{R}_n} + \varphi_{\mathbf{k}\boldsymbol{R}_j} - \varphi_{\mathbf{k}\boldsymbol{R}_j}\} \tag{2.2.5.40}$$

provided that $\mathbf{h}$ and $\boldsymbol{R}_n$ vary over the vectors and matrices for which (2.2.5.36) is compatible, $\mathbf{k}$ over the asymmetric region of the reciprocal space, and $\boldsymbol{R}_j$ over the rotation matrices in the space group. Formulae estimating $\varphi_{\mathbf{H}}$ *via* the second representation in all the space groups [all the base and cross magnitudes of the quintets (2.2.5.40) now constitute the *a priori* information] have been secured (Giacovazzo, 1978; Cascarano & Giacovazzo, 1983; Cascarano, Giacovazzo, Calabrese *et al.*, 1984). Such formulae contain, besides the contribution of order $N^{-1/2}$ provided by the first representation, a supplementary (not negligible) contribution of order $N^{-3/2}$ arising from quintets.

Denoting

$$E_1 = E_{\mathbf{H}},\ E_2 = E_{\mathbf{h}},\ E_3 = E_{\mathbf{k}},$$
$$E_{4,\,j} = E_{\mathbf{h}+\mathbf{k}\boldsymbol{R}_j},\ E_{5,\,j} = E_{\mathbf{H}+\mathbf{k}\boldsymbol{R}_j},$$

formulae (2.2.5.37), (2.2.5.38), (2.2.5.39) still hold provided that $\sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n}$ is replaced by

$$\sum_{\mathbf{h},\,n} G_{\mathbf{h},\,n} + \sum_{\mathbf{h},\,\mathbf{k},\,n}{}' \frac{|E_{\mathbf{H}}|}{2N^{3/2}} \frac{A_{\mathbf{h},\,\mathbf{k},\,n}}{1 + B_{\mathbf{h},\,\mathbf{k},\,n}},$$

where

$$A_{\mathbf{h},\,\mathbf{k},\,n} = \left[(2|E_2|^2 - 1)\varepsilon_3 \left(\sum_{\substack{\boldsymbol{R}_i = \boldsymbol{R}_j \\ \boldsymbol{R}_j + \boldsymbol{R}_i\boldsymbol{R}_n = 0}} \varepsilon_{4,\,i}\varepsilon_{5,\,j} + \sum_{\substack{\boldsymbol{R}_j = \boldsymbol{R}_i\boldsymbol{R}_n \\ \boldsymbol{R}_i = \boldsymbol{R}_j\boldsymbol{R}_n}} \varepsilon_{4,\,i}\varepsilon_{4,\,j} \right) - \frac{\varepsilon_3}{2} \sum_{j=1}^{m} \varepsilon_{4,\,j} - \tfrac{1}{2} \sum_{\substack{\boldsymbol{R}_j = \boldsymbol{R}_i \\ \boldsymbol{R}_j + \boldsymbol{R}_i\boldsymbol{R}_n = 0}} \varepsilon_{4,\,i}\varepsilon_{5,\,j} \right] \Big/ N,$$

$$B_{\mathbf{h},\,\mathbf{k},\,n} = \left[\varepsilon_1\varepsilon_3 \sum_{j=1}^{m} \varepsilon_{5,\,j} + \varepsilon_1 \sum_{\substack{\boldsymbol{R}_j = \boldsymbol{R}_i\boldsymbol{R}_n \\ \boldsymbol{R}_i = \boldsymbol{R}_j\boldsymbol{R}_n}} \varepsilon_{4,\,i}\varepsilon_{4,\,j} + \varepsilon_2\varepsilon_3 \sum_{j=1}^{m} \varepsilon_{4,\,j} + \varepsilon_2 \sum_{\substack{\boldsymbol{R}_j = \boldsymbol{R}_i \\ \boldsymbol{R}_j + \boldsymbol{R}_i\boldsymbol{R}_n = 0}} \varepsilon_{4,\,i}\varepsilon_{5,\,j} + \tfrac{1}{4}\varepsilon_1 H_4(E_2) \right] \Big/ (2N).$$

m is the number of symmetry operators and $H_4(E) = E^4 - 6E^2 + 3$ is the Hermite polynomial of order four.

$B_{\mathbf{h},\,\mathbf{k},\,n}$ is assumed to be zero if it is computed negative. The prime to the summation warns the reader that precautions have to be taken in order to avoid duplication in the contributions.

2.2.5.10. Formulae estimating two-phase structure seminvariants of the first rank

Two-phase s.s.'s of the first rank were first evaluated in some cs. space groups by the method of coincidence by Grant *et al.* (1957); the idea was extended to ncs. space groups by Debaerdemaeker & Woolfson (1972), and in a more general way by Giacovazzo (1977*e*,*f*).

The technique was based on the combination of the two triplets

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} \simeq \varphi_{\mathbf{h}_1+\mathbf{h}_2}$$
$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2\boldsymbol{R}} \simeq \varphi_{\mathbf{h}_1+\mathbf{h}_2\boldsymbol{R}},$$

which, subtracted from one another, give

$$\varphi_{\mathbf{h}_1+\mathbf{h}_2\boldsymbol{R}} - \varphi_{\mathbf{h}_1+\mathbf{h}_2} \simeq \varphi_{\mathbf{h}_2\boldsymbol{R}} - \varphi_{\mathbf{h}_2} \simeq -2\pi\mathbf{h}\cdot\mathbf{T}.$$

If all four $|E|$'s are sufficiently large, an estimate of the two-phase seminvariant $\varphi_{\mathbf{h}_1+\mathbf{h}_2\boldsymbol{R}} - \varphi_{\mathbf{h}_1+\mathbf{h}_2}$ is available.

Probability distributions valid in $P2_1$ according to the neighbourhood principle have been given by Hauptman & Green (1978). Finally, the theory of representations was combined by Giacovazzo (1979*a*) with the joint probability distribution method in order to estimate two-phase s.s.'s in all the space groups.

According to representation theory, the problem is that of evaluating $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$ *via* the special quartets (2.2.5.35*a*) and (2.2.5.35*b*). Thus, contributions of order N^{-1} will appear in the probabilistic formulae, which will be functions of the basis and of the cross magnitudes of the quartets (2.2.5.35) . Since more pairs of matrices $\boldsymbol{R}_\alpha$ and $\boldsymbol{R}_\beta$ can be compatible with (2.2.5.34), and for each pair $(\boldsymbol{R}_\alpha, \boldsymbol{R}_\beta)$ more pairs of vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ may satisfy (2.2.5.34), several quartets can in general be exploited for estimating Φ. The simplest case occurs in $P\bar{1}$ where the two quartets (2.2.5.35) suggest the calculation of the six-variate distribution function $(\mathbf{u}_1 = \mathbf{h}_1 + \mathbf{h}_2, \mathbf{u}_2 = \mathbf{h}_1 - \mathbf{h}_2)$

$$P(E_{\mathbf{h}_1}, E_{\mathbf{h}_2}, E_{\mathbf{h}_1+\mathbf{h}_2}, E_{\mathbf{h}_1-\mathbf{h}_2}, E_{2\mathbf{h}_1}, E_{2\mathbf{h}_2})$$

which leads to the probability formula

$$P^+ \simeq 0.5 + 0.5 \tanh\left(\frac{|E_{\mathbf{h}_1+\mathbf{h}_2} E_{\mathbf{h}_1-\mathbf{h}_2}|}{2N} \cdot \frac{A}{1+B}\right),$$

where P^+ is the probability that the product $E_{\mathbf{h}_1+\mathbf{h}_2} E_{\mathbf{h}_1-\mathbf{h}_2}$ is positive, and

$$\begin{aligned} A &= \varepsilon_{\mathbf{h}_1} + \varepsilon_{\mathbf{h}_2} + 2\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2} + \varepsilon_{\mathbf{h}_1}\varepsilon_{2\mathbf{h}_1} + \varepsilon_{\mathbf{h}_2}\varepsilon_{2\mathbf{h}_2} \\ B &= (\varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{u}_1} + \varepsilon_{\mathbf{h}_1}\varepsilon_{\mathbf{h}_2}\varepsilon_{\mathbf{u}_2} \\ &\quad + \varepsilon_{\mathbf{u}_1}\varepsilon_{\mathbf{u}_2}\varepsilon_{2\mathbf{h}_1} + \varepsilon_{\mathbf{u}_1}\varepsilon_{\mathbf{u}_2}\varepsilon_{2\mathbf{h}_2})/(2N). \end{aligned}$$

It may be seen that in favourable cases $P^+ < 0.5$.

For the sake of brevity, the probabilistic formulae for the general case are not given and the reader is referred to the original papers.

2.2.6. Direct methods in real and reciprocal space: Sayre's equation

The statistical treatment suggested by Wilson for scaling observed intensities corresponds, in direct space, to the origin peak of the Patterson function, so it is not surprising that a general correspondence exists between probabilistic formulation in reciprocal space and algebraic properties in direct space.

For a structure containing atoms which are fully resolved from one another, the operation of raising $\rho(\mathbf{r})$ to the nth power retains the condition of resolved atoms but changes the shape of each atom. Let

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} \rho_j(\mathbf{r} - \mathbf{r}_j),$$

where $\rho_j(\mathbf{r})$ is an atomic function and $\mathbf{r}_j$ is the coordinate of the 'centre' of the atom. Then the Fourier transform of the electron density can be written as

$$\begin{aligned} F_{\mathbf{h}} &= \sum_{j=1}^{N} \int_V \rho_j(\mathbf{r} - \mathbf{r}_j) \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\, \mathrm{d}V \\ &= \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j). \end{aligned} \tag{2.2.6.1}$$

If the atoms do not overlap

$$\rho^n(\mathbf{r}) = \left[\sum_{j=1}^{N} \rho_j(\mathbf{r} - \mathbf{r}_j)\right]^n \simeq \sum_{j=1}^{N} \rho_j^n(\mathbf{r} - \mathbf{r}_j)$$

and its Fourier transform gives

$$\begin{aligned} {}_nF_{\mathbf{h}} &= \int_V \rho^n(\mathbf{r}) \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\, \mathrm{d}V \\ &= \sum_{j=1}^{N} {}_nf_j \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j). \end{aligned} \tag{2.2.6.2}$$

${}_nf_j$ is the scattering factor for the jth peak of $\rho^n(\mathbf{r})$:

$${}_nf_j(\mathbf{h}) = \int_V \rho_j^n(\mathbf{r}) \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\, \mathrm{d}\mathbf{r}.$$

We now introduce the condition that all atoms are equal, so that $f_j \equiv f$ and ${}_nf_j \equiv {}_nf$ for any j. From (2.2.6.1) and (2.2.6.2) we may write

$$F_{\mathbf{h}} = \frac{f}{{}_nf} {}_nF_{\mathbf{h}} = \theta_n \, {}_nF_{\mathbf{h}}, \tag{2.2.6.3}$$

where θ_n is a function which corrects for the difference of shape of the atoms with electron distributions $\rho(\mathbf{r})$ and $\rho^n(\mathbf{r})$. Since

$$\begin{aligned} \rho^n(\mathbf{r}) &= \rho(\mathbf{r}) \ldots \rho(\mathbf{r}) \\ &= \frac{1}{V^n} \sum_{\substack{\mathbf{h}_1,\ldots,\mathbf{h}_n \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} \ldots F_{\mathbf{h}_n} \exp[-2\pi i(\mathbf{h}_1 + \ldots + \mathbf{h}_n)\cdot\mathbf{r}], \end{aligned}$$

the Fourier transform of both sides gives

$$\begin{aligned} {}_nF_{\mathbf{h}} &= \frac{1}{V^n} \sum_{\substack{\mathbf{h}_1,\ldots,\mathbf{h}_n \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} \ldots F_{\mathbf{h}_n} \int_V \exp[2\pi i(\mathbf{h} - \mathbf{h}_1 - \ldots - \mathbf{h}_n)\cdot\mathbf{r}]\, \mathrm{d}V \\ &= \frac{1}{V^{n-1}} \sum_{\substack{\mathbf{h}_1,\ldots,\mathbf{h}_{n-1} \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} F_{\mathbf{h}_2} \ldots F_{\mathbf{h}-\mathbf{h}_1-\mathbf{h}_2-\ldots-\mathbf{h}_{n-1}}, \end{aligned}$$

from which the following relation arises:

$$F_{\mathbf{h}} = \theta_n \frac{1}{V^{n-1}} \sum_{\substack{\mathbf{h}_1,\ldots,\mathbf{h}_{n-1} \\ -\infty}}^{+\infty} F_{\mathbf{h}_1} F_{\mathbf{h}_2} \ldots F_{\mathbf{h}-\mathbf{h}_1-\mathbf{h}_2-\ldots-\mathbf{h}_{n-1}}. \tag{2.2.6.4}$$

For $n = 2$, equation (2.2.6.4) reduces to Sayre's (1952) equation [but see also Hughes (1953)]

$$F_{\mathbf{h}} = \theta_2 \frac{1}{V} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}. \tag{2.2.6.5}$$

If the structure contains resolved isotropic atoms of two types, P and Q, it is impossible to find a factor θ_2 such that the relation $F_{\mathbf{h}} = \theta_2 \, {}_2F_{\mathbf{h}}$ holds, since this would imply values of θ_2 such that $({}_2f)_P = \theta_2(f)_P$ and $({}_2f)_Q = \theta_2(f)_Q$ simultaneously. However, the following relationship can be stated (Woolfson, 1958):

$$F_{\mathbf{h}} = \frac{A_s}{V} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} + \frac{B_s}{V^2} \sum_{\mathbf{k},\mathbf{l}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\mathbf{h}-\mathbf{k}-\mathbf{l}}, \tag{2.2.6.6}$$

where A_s and B_s are adjustable parameters of $(\sin\theta)/\lambda$. Equation (2.2.6.6) can easily be generalized to the case of structures containing resolved atoms of more than two types (von Eller, 1973).

Besides the algebraic properties of the electron density, Patterson methods also can be developed so that they provide phase indications. For example, it is possible to find the reciprocal counterpart of the function

$$P_n(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \int_V \rho(\mathbf{r})\rho(\mathbf{r}+\mathbf{u}_1)\ldots\rho(\mathbf{r}+\mathbf{u}_n)\, \mathrm{d}V. \tag{2.2.6.7}$$

For $n = 1$ the function (2.2.6.7) coincides with the usual Patterson function $P(\mathbf{u})$; for $n = 2$, (2.2.6.7) reduces to the double Patterson function $P_2(\mathbf{u}_1, \mathbf{u}_2)$ introduced by Sayre (1953). Expansion of $P_2(\mathbf{u}_1, \mathbf{u}_2)$ as a Fourier series yields

$$P_2(\mathbf{u}_1, \mathbf{u}_2) = \frac{1}{V^2} \sum_{\mathbf{h}_1, \mathbf{h}_2} E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3} \exp[-2\pi i(\mathbf{h}_1 \cdot \mathbf{u}_1 + \mathbf{h}_2 \cdot \mathbf{u}_2)]. \tag{2.2.6.8}$$

Vice versa, the value of a triplet invariant may be considered as the Fourier transform of the double Patterson.

Among the main results relating direct- and reciprocal-space properties it may be remembered:

(*a*) from the properties of $P_2(\mathbf{u}_1, \mathbf{u}_2)$ the following relationship may be obtained (Vaughan, 1958)

$$E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_1+\mathbf{h}_2} - N^{-3/2} \simeq A_1 \langle (|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_1+\mathbf{k}}|^2 - 1)(|E_{-\mathbf{h}_2+\mathbf{k}}|^2 - 1) \rangle_{\mathbf{k}} - B_1,$$

which is clearly related to (2.2.5.12);

(*b*) the zero points in the Patterson function provide information about the value of a triplet invariant (Anzenhofer & Hoppe, 1962; Allegra, 1979);

(*c*) the Hoppe sections (Hoppe, 1963) of the double Patterson provide useful information for determining the triplet signs (Krabbendam & Kroon, 1971; Simonov & Weissberg, 1970);

(*d*) one phase s.s.'s of the first rank can be estimated *via* the Fourier transform of single Harker sections of the Patterson (Ardito *et al.*, 1985), *i.e.*

$$F_{\mathbf{H}} \sim \frac{1}{L} \exp(2\pi i \mathbf{h} \cdot \mathbf{T}_n) \int_{HS(\mathbf{I}, \mathbf{C}_n)} P(\mathbf{u}) \exp(2\pi i \mathbf{h} \cdot \mathbf{u}) \, \mathrm{d}\mathbf{u}, \tag{2.2.6.9}$$

where (see Section 2.2.5.9) $\mathbf{H} = \mathbf{h}(\mathbf{I} - \mathbf{R}_n)$ is the s.s., $\mathbf{u}$ varies over the complete Harker section corresponding to the operator $\mathbf{C}_n$ [in symbols $HS(\mathbf{I}, \mathbf{C}_n)$] and L is a constant which takes into account the dimensionality of the Harker section.

If no spurious peak is on the Harker section, then (2.2.6.9) is an exact relationship. Owing to the finiteness of experimental data and to the presence of spurious peaks, (2.2.6.9) cannot be considered in practice an exact relation: it works better when heavy atoms are in the chemical formula.

More recently (Cascarano, Giacovazzo, Luić *et al.*, 1987), a special least-squares procedure has been proposed for discriminating spurious peaks among those lying on Harker sections and for improving positional and thermal parameters of heavy atoms.

(*e*) translation and rotation functions (see Chapter 2.3), when defined in direct space, always have their counterpart in reciprocal space.

2.2.7. Scheme of procedure for phase determination: the small-molecule case

A traditional procedure for phase assignment may be schematically presented as follows:

Stage 1: *Normalization of s.f.'s.* See Section 2.2.4.

Stage 2: *(Possible) estimation of one-phase s.s.'s.* The computing program recognizes the one-phase s.s.'s and applies the proper formulae (see Section 2.2.5.9).

Each phase is associated with a reliability value, to allow the user to regard as known only those phases with reliability higher than a given threshold.

Stage 3: *Search of the triplets.* The reflections are listed for decreasing $|E|$ values and, related to each $|E|$ value, all possible triplets are reported (this is the so-called $\sum_2$ list). The value $G = 2|E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}|/\sqrt{N}$ is associated with every triplet for an evaluation of its efficiency. Usually reflections with $|E| < E_s$ (E_s may range from 1.2 to 1.6) are omitted from this stage onward.

Stage 4: *Definition of the origin and enantiomorph.* This stage is carried out according to the theory developed in Section 2.2.3. Phases chosen for defining the origin and enantiomorph, one-phase seminvariants estimated at stage 2, and symbolic phases described at stage 5 are the only phases known at the beginning of the phasing procedure. This set of phases is conventionally referred to as the *starting set*, from which iterative application of the tangent formula will derive new phase estimates.

Stage 5: *Assignment of one or more (symbolic or numerical) phases.* In complex structures the number of phases assigned for fixing the origin and the enantiomorph may be inadequate as a basis for further phase determination. Furthermore, only a few one-phase s.s.'s can be determined with sufficient reliability to make them qualify as members of the starting set. Symbolic phases may then be associated with some (generally from 1 to 6) high-modulus reflections (symbolic addition procedures). Iterative application of triplet relations leads to the determination of other phases which, in part, will remain expressed by symbols (Karle & Karle, 1966).

In other procedures (multisolution procedures) each symbol is assigned four phase values in turn: $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. If p symbols are used, in at least one of the possible 4^p solutions each symbolic phase has unit probability of being within 45° of its true value, with a mean error of 22.5°.

To find a good starting set a *convergence method* (Germain *et al.*, 1970) is used according to which: (*a*)

$$\langle \alpha_{\mathbf{h}} \rangle = \sum_j G_j I_1(G_j)/I_0(G_j)$$

is calculated for all reflections (j runs over the set of triplets containing $\mathbf{h}$); (*b*) the reflection is found with smallest $\langle \alpha \rangle$ not already in the starting set; it is retained to define the origin if the origin cannot be defined without it; (*c*) the reflection is eliminated if it is not used for origin definition. Its $\langle \alpha \rangle$ is recorded and $\langle \alpha \rangle$ values for other reflections are updated; (*d*) the cycle is repeated from (*b*) until all reflections are eliminated; (*e*) the reflections with the smallest $\langle \alpha \rangle$ at the time of elimination go into the starting set; (*f*) the cycle from (*a*) is repeated until all reflections have been chosen.

Stage 6: *Application of tangent formula.* Phases are determined in reverse order of elimination in the convergence procedure. In order to ensure that poorly determined phases $\varphi_{\mathbf{k}_j}$ and $\varphi_{\mathbf{h}-\mathbf{k}_j}$ have little effect in the determination of other phases a weighted tangent formula is normally used (Germain *et al.*, 1971):

$$\tan \varphi_{\mathbf{h}} = \frac{\sum_j w_{\mathbf{k}_j} w_{\mathbf{h}-\mathbf{k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j w_{\mathbf{k}_j} w_{\mathbf{h}-\mathbf{k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}, \tag{2.2.7.1}$$

where

$$w_{\mathbf{h}} = \min(0.2\alpha, 1).$$

Once a large number of contributions are available in (2.2.7.1) for a given $\varphi_{\mathbf{h}}$, then the value of $\alpha_{\mathbf{h}}$ quickly becomes greater than 5, and so assigns an unrealistic unitary weight to $\varphi_{\mathbf{h}}$. In this respect a different weighting scheme may be proposed (Hull & Irwin, 1978) according to which

$$w = \psi \exp(-x^2) \int_0^x \exp(t^2) \, \mathrm{d}t, \tag{2.2.7.2}$$

where $x = \alpha/\langle \alpha \rangle$ and $\psi = 1.8585$ is a constant chosen so that $w = 1$ when $x = 1$. Except for ψ, the right-hand side of (2.2.7.2) is the Dawson integral which assumes its maximum value at $x = 1$ (see Fig. 2.2.7.1): when $\alpha > \langle \alpha \rangle$ or $\alpha < \langle \alpha \rangle$ then $w < 1$ and so the agreement between α and $\langle \alpha \rangle$ is promoted.

Alternative weighting schemes for the tangent formula are frequently used [for example, see Debaerdemaeker *et al.* (1985)]. In one (Giacovazzo, 1979*b*), the values $\alpha_{\mathbf{k}_j}$ and $\alpha_{\mathbf{h}-\mathbf{k}_j}$ (which are

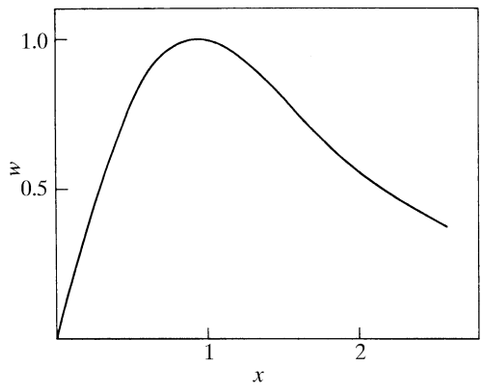


Fig. 2.2.7.1. The form of w as given by (2.2.7.2).

usually available in direct procedures) are considered as additional *a priori* information so that (2.2.7.1) may be replaced by

$$\tan\varphi_{\mathbf{h}} \simeq \frac{\sum_j \beta_j \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j \beta_j \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}, \tag{2.2.7.3}$$

where β_j is the solution of the equation

$$D_1(\beta_j) = D_1(G_j) D_1(\alpha_{\mathbf{k}_j}) D_1(\alpha_{\mathbf{h}-\mathbf{k}_j}). \tag{2.2.7.4}$$

In (2.2.7.4),

$$G_j = 2|E_{\mathbf{h}} E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}|\sqrt{N}$$

or the corresponding second representation parameter, and $D_1(x) = I_1(x)/I_0(x)$ is the ratio of two modified Bessel functions.

In order to promote (in accordance with the aims of Hull and Irwin) the agreement between α and $\langle\alpha\rangle$, the distribution of α may be used (Cascarano, Giacovazzo, Burla *et al.*, 1984; Burla *et al.*, 1987); in particular, the first two moments of the distribution: accordingly,

$$w = \left\{\exp\left[\frac{-(\alpha - \langle\alpha\rangle)^2}{2\sigma_\alpha^2}\right]\right\}^{1/3}$$

may be used, where σ_α^2 is the estimated variance of α.

Stage 7: *Figures of merit.* The correct solution is found among several by means of figures of merit (FOMs) which are expected to be extreme for the correct solution. Largely used are (Germain *et al.*, 1970)

$$(a) \qquad \text{ABSFOM} = \sum_{\mathbf{h}} \alpha_{\mathbf{h}} / \sum_{\mathbf{h}} \langle\alpha_{\mathbf{h}}\rangle,$$

which is expected to be unity for the correct solution.

$$(b) \qquad \text{PSI0} = \frac{\sum_{\mathbf{h}} \left|\sum_{\mathbf{k}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}\right|}{\sum_{\mathbf{h}} \left(\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}|^2\right)^{1/2}}.$$

The summation over $\mathbf{k}$ includes (Cochran & Douglas, 1957) the strong $|E|$'s for which phases have been determined, and indices $\mathbf{h}$ correspond to very small $|E_{\mathbf{h}}|$. Minimal values of PSI0 (≤ 1.20) are expected to be associated with the correct solution.

$$(c) \qquad R_\alpha = \frac{\sum_{\mathbf{h}} |\alpha_{\mathbf{h}} - \langle\alpha_{\mathbf{h}}\rangle|}{\sum_{\mathbf{h}} \langle\alpha_{\mathbf{h}}\rangle}.$$

That is, the Karle & Karle (1966) residual between the actual and the estimated α's. After scaling of $\alpha_{\mathbf{h}}$ on $\langle\alpha_{\mathbf{h}}\rangle$ the correct solution should be characterized by the smallest R_α values.

$$(d) \qquad \text{NQEST} = \sum_j G_j \cos\Phi_j,$$

where G is defined by (2.2.5.21) and

$$\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{l}} - \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}}$$

are quartet invariants characterized by large basis magnitudes and small cross magnitudes (De Titta *et al.*, 1975; Giacovazzo, 1976). Since G is expected to be negative as well as $\cos\Phi$, the value of NQEST is expected to be positive and a maximum for the correct solution.

Figures of merit are then combined as

$$\begin{aligned}\text{CFOM} = {} & w_1 \frac{\text{ABSFOM} - \text{ABSFOM}_{\min}}{\text{ABSFOM}_{\max} - \text{ABSFOM}_{\min}} \\ & + w_2 \frac{\text{PSI0}_{\max} - \text{PSI0}}{\text{PSI0}_{\max} - \text{PSI0}_{\min}} \\ & + w_3 \frac{R_{\alpha_{\max}} - R_\alpha}{R_{\alpha_{\max}} - R_{\alpha_{\min}}} \\ & + w_4 \frac{\text{NQEST} - \text{NQEST}_{\min}}{\text{NQEST}_{\max} - \text{NQEST}_{\min}},\end{aligned}$$

where w_i are empirical weights proportional to the confidence of the user in the various FOMs.

Different FOMs are often used by some authors in combination with those described above: for example, enantiomorph triplets and quartets are supplementary FOMs (Van der Putten & Schenk, 1977; Cascarano, Giacovazzo & Viterbo, 1987).

Different schemes of calculating and combining FOMs are also used: one scheme (Cascarano, Giacovazzo & Viterbo, 1987) uses

$$(a1) \qquad \text{CPHASE} = \frac{\sum w_j G_j \cos(\Phi_j - \theta_j) + w_j G_j \cos\Phi_j}{\sum_{\text{s.i.+s.s.}} w_j G_j D_1(G_j)},$$

where the first summation in the numerator extends over symmetry-restricted one-phase and two-phase s.s.'s (see Sections 2.2.5.9 and 2.2.5.10), and the second summation in the numerator extends over negative triplets estimated *via* the second representation formula [equation (2.2.5.13)] and over negative quartets. The value of CPHASE is expected to be close to unity for the correct solution.

(*a*2) $\alpha_{\mathbf{h}}$ for strong triplets and $E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}$ contributions for PSI0 triplets may be considered random variables: the agreements between their actual and their expected distributions are considered as criteria for identifying the correct solution.

(*a*3) correlation among some FOMs is taken into account.

According to this scheme, each FOM (as well as the CFOM) is expected to be unity for the correct solution. Thus one or more figures are available which constitute a sort of criterion (on an absolute scale) concerning the correctness of the various solutions: FOMs (and CFOM) $\simeq 1$ probably denote correct solutions, CFOMs $\ll 1$ should indicate incorrect solutions.

Stage 8: *Interpretation of E maps.* This is carried out in up to four stages (Koch, 1974; Main & Hull, 1978; Declercq *et al.*, 1973):

(*a*) peak search;

(*b*) separation of peaks into potentially bonded clusters;

(*c*) application of stereochemical criteria to identify possible molecular fragments;

(*d*) comparison of the fragments with the expected molecular structure.

2.2.8. Other multisolution methods applied to small molecules

In very complex structures a large initial set of known phases seems to be a basic requirement for a structure to be determined.

Table 2.2.8.1. *Magic-integer sequences for small numbers of phases (n) together with the number of sets produced and the root-mean-square error in the phases*

n	Sequence								No. of sets	R.m.s. error (°)
1	1								4	26
2	2	3							12	29
3	3	4	5						20	37
4	5	7	8	9					32	42
5	8	11	13	14	15				50	45
6	13	18	21	23	24	25			80	47
7	21	29	34	37	39	40	41		128	48
8	34	47	55	60	63	65	66	67	206	49

This aim can be achieved, for example, by introducing a large number of permutable phases into the initial set. However, the introduction of every new symbol implies a fourfold increase in computing time, which, even in fast computers, quickly leads to computing-time limitations. On the other hand, a relatively large starting set is not in itself enough to ensure a successful structure determination. This is the case, for example, when the triplet invariants used in the initial steps differ significantly from zero. New strategies have therefore been devised to solve more complex structures.

(1) *Magic-integer methods*

In the classical procedure described in Section 2.2.7, the unknown phases in the starting set are assigned all combinations of the values $\pm\pi/4$, $\pm 3\pi/4$. For n unknown phases in the starting set, 4^n sets of phases arise by quadrant permutation; this is a number that increases very rapidly with n. According to White & Woolfson (1975), phases can be represented for a sequence of n integers by the equations

$$\varphi_i = m_i x \pmod{2\pi}, \quad i = 1, \ldots, n. \tag{2.2.8.1}$$

The set of equations can be regarded as the parametric equation of a straight line in n-dimensional phase space. The nature and size of errors connected with magic-integer representations have been investigated by Main (1977) who also gave a recipe for deriving magic-integer sequences which minimize the r.m.s. errors in the represented phases (see Table 2.2.8.1). To assign a phase value, the variable x in equation (2.2.8.1) is given a series of values at equal intervals in the range $0 < x < 2\pi$. The enantiomorph is defined by exploring only the appropriate half of the n-dimensional space.

A different way of using the magic-integer method (Declercq *et al.*, 1975) is the *primary–secondary P–S method* which may be described schematically in the following way:

(a) Origin- and enantiomorph-fixing phases are chosen and some one-phase s.s.'s are estimated.

(b) Nine phases [this is only an example: very long magic-integer sequences may be used to represent primary phases (Hull *et al.*, 1981; Debaerdemaeker & Woolfson, 1983)] are represented with the approximated relationships:

$$\begin{cases} \varphi_{i_1} = 3x \\ \varphi_{i_2} = 4x \\ \varphi_{i_3} = 5x \end{cases} \qquad \begin{cases} \varphi_{j_1} = 3y \\ \varphi_{j_2} = 4y \\ \varphi_{j_3} = 5y \end{cases} \qquad \begin{cases} \varphi_{p_1} = 3z \\ \varphi_{p_2} = 4z \\ \varphi_{p_3} = 5z. \end{cases}$$

Phases in (a) and (b) consistitute the *primary set.*

(c) The phases in the *secondary set* are those defined through $\sum_2$ relationships involving pairs of phases from the primary set: they, too, can be expressed in magic-integer form.

(d) All the triplets that link together the phases in the combined primary and secondary set are now found, other than triplets used to obtain secondary reflections from the primary ones. The general algebraic form of these triplets will be

$$m_1 x + m_2 y + m_3 z + b \equiv 0 \pmod 1,$$

where b is a phase constant which arises from symmetry translation. It may be expected that the 'best' value of the unknown x, y, z corresponds to a maximum of the function

$$\psi(x, y, z) = \sum |E_1 E_2 E_3| \cos 2\pi(m_1 x + m_2 y + m_3 z + b),$$

with $0 \le x, y, z < 1$. It should be noticed that ψ is a Fourier summation which can easily be evaluated. In fact, ψ is essentially a figure of merit for a large number of phases evaluated in terms of a small number of magic-integer variables and gives a measure of the internal consistency of $\sum_2$ relationships. The ψ map generally presents several peaks and therefore can provide several solutions for the variables.

(2) *The random-start method*

These are procedures which try to solve crystal structures by starting from random initial phases (Baggio *et al.*, 1978; Yao, 1981). They may be so described:

(a) A number of reflections (say NUM $\sim$ 100 or larger) at the bottom of the CONVERGE map are selected. These, and the relationships which link them, form the system for which trial phases will be found.

(b) A pseudo-random number generator is used to generate M sets of NUM random phases. Each of the M sets is refined and extended by the tangent formula or similar methods.

(3) *Accurate calculation of s.i.'s and s.s.'s with* 1, 2, 3, 4, ..., n *phases*

Having a large set of good phase relationships allows one to overcome difficulties in the early stages and in the refinement process of the phasing procedure. Accurate estimates of s.i.'s and s.s.'s may be achieved by the application of techniques such as the representation method or the neighbourhood principle (Hauptman, 1975; Giacovazzo, 1977*a*, 1980*b*). So far, second-representation formulae are available for triplets and one-phase seminvariants; in particular, reliably estimated negative triplets can be recognized, which is of great help in the phasing process (Cascarano, Giacovazzo, Camalli *et al.*, 1984). Estimation of higher-order s.s.'s with upper representations or upper neighbourhoods is rather difficult, both because the procedures are time consuming and because the efficiency of the present joint probability distribution techniques deteriorates with complexity. However, further progress can be expected in the field.

(4) *Modified tangent formulae and least-squares determination and refinement of phases*

The problem of deriving the individual phase angles from triplet relationships is greatly overdetermined: indeed the number of triplets, in fact, greatly exceeds the number of phases so that any $\varphi_{\mathbf{h}}$ may be determined by a least-squares approach (Hauptman *et al.*, 1969). The function to be minimized may be

$$M = \frac{\sum_{\mathbf{k}} w_{\mathbf{k}} [\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}) - C_{\mathbf{k}}]^2}{\sum w_{\mathbf{k}}},$$

where $C_{\mathbf{k}}$ is the estimate of the cosine obtained by probabilistic or other methods.

Effective least-squares procedures based on linear equations (Debaerdemaeker & Woolfson, 1983; Woolfson, 1977) can also be used. A triplet relationship is usually represented by

$$(\varphi_p \pm \varphi_q \pm \varphi_r + b) \approx 0 \pmod{2\pi}, \tag{2.2.8.2}$$

where b is a factor arising from translational symmetry. If (2.2.8.2) is expressed in cycles and suitably weighted, then it may be written as

$$w(\varphi_p \pm \varphi_q \pm \varphi_r + b) = wn,$$

where n is some integer. If the integers were known then the equation would appear (in matrix notation) as

$$\boldsymbol{A}\boldsymbol{\Phi} = \boldsymbol{C}, \tag{2.2.8.3}$$

giving the least-squares solution

$$\boldsymbol{\Phi} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{C}. \tag{2.2.8.4}$$

When approximate phases are available, the nearest integers may be found and equations (2.2.8.3) and (2.2.8.4) constitute the basis for further refinement.

Modified tangent procedures are also used, such as (Sint & Schenk, 1975; Busetta, 1976)

$$\tan\varphi_{\mathbf{h}} \simeq \frac{\sum_j G_{\mathbf{h},\mathbf{k}_j}\sin(\varphi_{\mathbf{k}_j}+\varphi_{\mathbf{h}-\mathbf{k}_j}-\Delta_j)}{\sum G_{\mathbf{h},\mathbf{k}_j}\cos(\varphi_{\mathbf{k}_j}+\varphi_{\mathbf{h}-\mathbf{k}_j}-\Delta_j)},$$

where Δ_j is an estimate for the triplet phase sum $(\varphi_{\mathbf{h}}-\varphi_{\mathbf{k}_j}-\varphi_{\mathbf{h}-\mathbf{k}_j})$.

(5) *Techniques based on the positivity of Karle–Hauptman determinants*

(The main formulae have been briefly described in Section 2.2.5.7.) The maximum determinant rule has been applied to solve small structures (de Rango, 1969; Vermin & de Graaff, 1978) *via* determinants of small order. It has, however, been found that their use (Taylor *et al.*, 1978) is not of sufficient power to justify the larger amount of computing time required by the technique as compared to that required by the tangent formula.

(6) *Tangent techniques using simultaneously triplets, quartets, ...*

The availability of a large number of phase relationships, in particular during the first stages of a direct procedure, makes the phasing process easier. However, quartets are sums of two triplets with a common reflection. If the phase of this reflection (and/or of the other cross terms) is known then the quartet probability formulae described in Section 2.2.5.5 cannot hold. Similar considerations may be made for quintet relationships. Thus triplet, quartet and quintet formulae described in the preceding paragraphs, if used without modifications, will certainly introduce systematic errors in the tangent refinement process.

A method which takes into account correlation between triplets and quartets has been described (Giacovazzo, 1980*c*) [see also Freer & Gilmore (1980) for a first application], according to which

$$\tan\varphi_{\mathbf{h}} \simeq \frac{\sum_{\mathbf{k}} G\sin(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}}) - \sum_{\mathbf{k},\mathbf{l}} G'\sin(\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})}{\sum_{\mathbf{k}} G\cos(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}}) - \sum_{\mathbf{k},\mathbf{l}} G'\cos(\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})},$$

where G' takes into account both the magnitudes of the cross terms of the quartet and the fact that their phases may be known.

(7) *Integration of Patterson techniques and direct methods (Egert & Sheldrick, 1985) [see also Egert (1983, and references therein)]*

A fragment of known geometry is oriented in the unit cell by real-space Patterson rotation search (see Chapter 2.3) and its position is found by application of a translation function (see Section 2.2.5.4 and Chapter 2.3) or by maximizing the weighted sum of the cosines of a small number of strong translation-sensitive triple phase invariants, starting from random positions. Suitable FOMs rank the most reliable solutions.

(8) *Maximum entropy methods*

A common starting point for all direct methods is a stochastic process according to which crystal structures are thought of as being generated by randomly placing atoms in the asymmetric unit of the unit cell according to some *a priori* distribution. A non-uniform prior distribution of atoms $p(\mathbf{r})$ gives rise to a source of random atomic positions with entropy (Jaynes, 1957)

$$H(p) = -\int_V p(\mathbf{r})\log p(\mathbf{r})\,\mathrm{d}\mathbf{r}.$$

The maximum value $H_{\max} = \log V$ is reached for a uniform prior $p(\mathbf{r}) = 1/V$.

The strength of the restrictions introduced by $p(\mathbf{r})$ is not measured by $H(p)$ but by $H(p) - H_{\max}$, given by

$$H(p) - H_{\max} = -\int_V p(\mathbf{r})\log[\,p(\mathbf{r})/m(\mathbf{r})]\,\mathrm{d}\mathbf{r},$$

where $m(\mathbf{r}) = 1/V$. Accordingly, if a prior prejudice $m(\mathbf{r})$ exists, which maximizes H, the revised relative entropy is

$$S(p) = -\int_V p(\mathbf{r})\log[\,p(\mathbf{r})/m(\mathbf{r})]\,\mathrm{d}\mathbf{r}.$$

The maximization problem was solved by Jaynes (1957). If $G_j(p)$ are linear constraint functionals defined by given constraint functions $C_j(\mathbf{r})$ and constraint values c_j, *i.e.*

$$G_j(p) = \int_V p(\mathbf{r})C_j(\mathbf{r})\,\mathrm{d}\mathbf{r} = c_j,$$

the most unbiased probability density $p(\mathbf{r})$ under prior prejudice $m(\mathbf{r})$ is obtained by maximizing the entropy of $p(\mathbf{r})$ relative to $m(\mathbf{r})$. A standard variational technique suggests that the constrained maximization is equivalent to the unconstrained maximization of the functional

$$S(p) + \sum_j \lambda_j G_j(p),$$

where the λ_j's are Lagrange multipliers whose values can be determined from the constraints.

Such a technique has been applied to the problem of finding good electron-density maps in different ways by various authors (Wilkins *et al.*, 1983; Bricogne, 1984; Navaza, 1985; Navaza *et al.*, 1983).

Maximum entropy methods are strictly connected with traditional direct methods: in particular it has been shown that:

(*a*) the maximum determinant rule (see Section 2.2.5.7) is strictly connected (Britten & Collins, 1982; Piro, 1983; Narayan & Nityananda, 1982; Bricogne, 1984);

(*b*) the construction of conditional probability distributions of structure factors amounts precisely to a reciprocal-space evaluation of the entropy functional $S(p)$ (Bricogne, 1984).

Maximum entropy methods are under strong development: important contributions can be expected in the near future even if a multipurpose robust program has not yet been written.

2.2.9. Some references to direct-methods packages: the small-molecule case

Some references for direct-methods packages are given below. Other useful packages using symbolic addition or multisolution procedures do exist but are not well documented.

CRUNCH: Gelder, R. de, de Graaff, R. A. G. & Schenk, H. (1993). *Automatic determination of crystal structures using Karle–Hauptman matrices. Acta Cryst.* A**49**, 287–293.

DIRDIF: Beurskens, P. T., Beurskens G., de Gelder, R., Garcia-Granda, S., Gould, R. O., Israel, R. & Smits, J. M. M. (1999). *The DIRDIF-99 program system*. Crystallography Laboratory, University of Nijmegen, The Netherlands.

MITHRIL: Gilmore, C. J. (1984). *MITHRIL. An integrated direct-methods computer program. J. Appl. Cryst.* **17**, 42–46.

MULTAN88: Main, P., Fiske, S. J., Germain, G., Hull, S. E., Declercq, J.-P., Lessinger, L. & Woolfson, M. M. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

PATSEE: Egert, E. & Sheldrick, G. M. (1985). *Search for a fragment of known geometry by integrated Patterson and direct methods. Acta Cryst.* A**41**, 262–268.

SAPI: Fan, H.-F. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

SnB: Weeks, C. M. & Miller, R. (1999). *The design and implementation of SnB version 2.0. J. Appl. Cryst.* **32**, 120–124.

SHELX97 and *SHELXS*: Sheldrick, G. M. (2000). *The SHELX home page*. http://shelx.uni-ac.gwdg.de/SHELX/.

SHELXD: Sheldrick, G. M. (1998). *SHELX: applications to macromolecules*. In *Direct methods for solving macromolecular structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.

SIR97: Altomare, A., Burla, M. C., Camalli, M., Cascarano, G. L., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Spagna, R. (1999). *SIR97: a new tool for crystal structure determination and refinement. J. Appl. Cryst.* **32**, 115–119.

SIR2004: Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. (2005). *SIR2004: an improved tool for crystal structure determination and refinement. J. Appl. Cryst.* **38**, 381–388.

XTAL3.6.1: Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (1999). *Xtal3.6 crystallographic software.* http://xtal.sourceforge.net/.

2.2.10. Direct methods in macromolecular crystallography

2.2.10.1. Introduction

The smallest protein molecules contain about 400 non-hydrogen atoms, so they cannot be solved *ab initio* by the algorithms specified in Sections 2.2.7 and 2.2.8. However, traditional direct methods are applied for:

(*a*) improvement of the accuracy of the available phases (refinement process);

(*b*) extension of phases from lower to higher resolution (phase-extension process).

The application of standard tangent techniques to (*a*) and (*b*) has not been found to be very satisfactory (Coulter & Dewar, 1971; Hendrickson *et al.*, 1973; Weinzierl *et al.*, 1969). Tangent methods, in fact, require atomicity and non-negativity of the electron density. Both these properties are not satisfied if data do not extend to atomic resolution ($d > 1.2$ Å). Because of series termination and other errors the electron-density map at $d > 1.2$ Å presents large negative regions which will appear as false peaks in the squared structure. However, tangent methods use only a part of the information given by the Sayre equation (2.2.6.5). In fact, (2.2.6.5) express two equations relating the radial and angular parts of the two sides, so obtaining a large degree of overdetermination of the phases. To achieve this Sayre (1972) [see also Sayre & Toupin (1975)] suggested minimizing (2.2.10.1) by least squares as a function of the phases:

$$\sum_{\mathbf{h}} \left| a_{\mathbf{h}} F_{\mathbf{h}} - \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} \right|^2 . \qquad (2.2.10.1)$$

Even if tests on rubredoxin (extensions of phases from 2.5 to 1.5 Å resolution) and insulin (Cutfield *et al.*, 1975) (from 1.9 to 1.5 Å resolution) were successful, the limitations of the method are its high cost and, especially, the higher efficiency of the least-squares method. Equivalent considerations hold for the application of determinantal methods to proteins [see Podjarny *et al.* (1981); de Rango *et al.* (1985) and literature cited therein].

A question now arises: why is the tangent formula unable to solve protein structures? Fan *et al.* (1991) considered the question from a first-principle approach and concluded that:

(1) the triplet phase probability distribution is very flat for proteins (N is very large) and close to the uniform distribution;

(2) low-resolution data create additional problems for direct methods since the number of available phase relationships per reflection is small.

Sheldrick (1990) suggested that direct methods are not expected to succeed if fewer than half of the reflections in the range 1.1–1.2 Å are observed with $|F| > 4\sigma(|F|)$ (a condition seldom satisfied by protein data).

The most complete analysis of the problem has been made by Giacovazzo, Guagliardi *et al.* (1994). They observed that the expected value of α (see Section 2.2.7) suggested by the tangent formula for proteins is comparable with the variance of the α parameter. In other words, for proteins the signal determining the phase is comparable with the noise, and therefore the phase indication is expected to be unreliable.

Quite relevant results have recently been obtained by integrating direct methods with some additional experimental information. In particular, we will describe the combination of direct methods with:

(*a*) direct-space techniques for the *ab initio* crystal structure solution of proteins;

(*b*) isomorphous-replacement (SIR–MIR) techniques;

(*c*) anomalous-dispersion (SAD–MAD) techniques;

(*d*) molecular replacement.

Point (*d*) will not be treated here, as it is described extensively in *IT* F, Part 13.

2.2.10.2. Ab initio crystal structure solution of proteins

Ab initio techniques do not require prior information of any atomic positions. The recent tremendous increase in computing speed led to direct methods evolving towards the rapid development of multisolution techniques. The new algorithms of the program *Shake-and-Bake* (Weeks *et al.*, 1994; Weeks & Miller, 1999; Hauptman *et al.*, 1999) allowed an impressive extension of the structural complexity amenable to direct phasing. In particular we mention: (*a*) the minimal principle (De Titta *et al.*, 1994), according to which the phase problem is considered as a constrained global optimization problem; (*b*) the refinement procedure, which alternately uses direct- and reciprocal-space techniques; and (*c*) the parameter-shift optimization technique (Bhuiya & Stanley, 1963), which aims at reducing the value of the minimal function (Hauptman, 1991; De Titta *et al.*, 1994). An effective variant of *Shake-and Bake* is *SHELXD* (Sheldrick, 1998) which cyclically alternates tangent refinement in reciprocal space with peak-list optimisation procedures in real space (Sheldrick & Gould, 1995). Detailed information on these programs is available in *IT* F (2001), Part 16.

A different approach is used by *ACORN* (Foadi *et al.*, 2000), which first locates a small fragment of the molecule (eventually by molecular-replacement techniques) to obtain a useful nonrandom starting set of phases, and then refines them by means of solvent-flattening techniques.

The program *SIR2004* (Burla *et al.*, 2005) uses the tangent formula as well as automatic Patterson techniques to obtain a first imperfect structural model; then direct-space techniques are used to refine the model. The Patterson approach is based on the use of the superposition minimum function (Buerger, 1959; Richardson & Jacobson, 1987; Sheldrick, 1992; Pavelcík, 1988; Pavelcík *et al.*, 1992; Burla *et al.*, 2004). It may be worth noting that even this approach is of multisolution type: up to 20 trial solutions are provided by using as pivots the highest maxima in the superposition minimum function.

It is today possible to solve structures up to 2500 non-hydrogen atoms in the asymmetric unit provided data at atomic (about 1 Å) resolution are available. Proteins with data at quasi-atomic resolution (say up to 1.5–1.6 Å) can also be solved, but with greater difficulties (Burla *et al.*, 2005). A simple evaluation of the potential of the *ab initio* techniques suggests that the structural complexity range and the resolution limits amenable to the *ab*

initio approach could be larger in the near future. The approach will profit by general technical advances like the increasing speed of computers and by the greater efficiency of informatic tools (*e.g.* faster Fourier-transform techniques). It could also profit from new specific crystallographic algorithms (for example, Oszlányi & Süto, 2004). It is of particular interest that extrapolating moduli and phases of nonmeasured reflections beyond the experimental resolution limit makes the *ab initio* phasing process more efficient, and leads to crystal structure solution even in cases in which the standard programs do not succeed (Caliandro *et al.*, 2005*a*,*b*). Moreover, the use of the extrapolated values improves the quality of the final electron-density maps and makes it easier to recognize the correct one among several trial structures.

2.2.10.3. Integration of direct methods with isomorphous replacement techniques

SIR–MIR cases are characterized by a situation in which there is one native protein and one or more heavy-atom substructures. In this situation the phasing procedure may be a two-step process: in the first stage the heavy-atom positions are identified by Patterson techniques (Rossmann, 1961; Okaya *et al.*, 1955) or by direct methods (Mukherjee *et al.*, 1989). In the second step the protein phases are estimated by exploiting the substructure information. Direct methods are able to contribute to both steps (see Sections 2.2.10.5 and 2.2.10.6). In Section 2.2.10.4 we show that direct methods are also able to suggest alternative one-step procedures by estimating structure invariants from isomorphous data.

2.2.10.4. SIR–MIR case: one-step procedures

The theoretical basis was established by Hauptman (1982*a*): his primary interest was to establish the two-phase and three-phase structure invariants by exploiting the experimental information provided by isomorphous data. The protein phases could be directly assigned *via* a tangent procedure.

Let us denote the modulus of the isomorphous difference as

$$\Delta F = |F_{\rm d}| - |F_{\rm p}|$$

where the subscripts d and p denote the derivative and the protein, respectively.

Denote also by f_j and g_j atomic scattering factors for the atom labelled j in a pair of isomorphous structures, and let $E_{\mathbf{h}}$ and $G_{\mathbf{h}}$ denote corresponding normalized structure factors. Then

$$E_{\mathbf{h}} = |E_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) = \alpha_{20}^{-1/2} \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j),$$

$$G_{\mathbf{h}} = |G_{\mathbf{h}}| \exp(i\psi_{\mathbf{h}}) = \alpha_{02}^{-1/2} \sum_{j=1}^{N} g_j \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_j),$$

where

$$\alpha_{mn} = \sum_{j=1}^{N} f_j^m g_j^n.$$

The conditional probability of the two-phase structure invariant $\Phi = \varphi_{\mathbf{h}} - \psi_{\mathbf{h}}$ given $|E_{\mathbf{h}}|$ and $|G_{\mathbf{h}}|$ is (Hauptman, 1982*a*)

$$P(\Phi\big||E|, |G|) \simeq [2\pi I_0(Q)]^{-1} \exp(Q\cos\Phi),$$

where

$$Q = |EG|[2\alpha/(1-\alpha^2)],$$
$$\alpha = \alpha_{11}/(\alpha_{20}^{1/2}\alpha_{02}^{1/2}).$$

Three-phase structure invariants were evaluated by considering that eight invariants exist for a given triple of indices **h**, **k**, **l** ($\mathbf{h}+\mathbf{k}+\mathbf{l}=0$):

$$\Phi_1 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_2 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \psi_{\mathbf{l}}$$
$$\Phi_3 = \varphi_{\mathbf{h}} + \psi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_4 = \psi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}$$
$$\Phi_5 = \varphi_{\mathbf{h}} + \psi_{\mathbf{k}} + \psi_{\mathbf{l}} \qquad \Phi_6 = \psi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \psi_{\mathbf{l}}$$
$$\Phi_7 = \psi_{\mathbf{h}} + \psi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_8 = \psi_{\mathbf{h}} + \psi_{\mathbf{k}} + \psi_{\mathbf{l}}.$$

So, for the estimation of any Φ_j, the joint probability distribution

$$P(E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{l}}, G_{\mathbf{h}}, G_{\mathbf{k}}, G_{\mathbf{l}})$$

has to be studied, from which eight conditional probability densities can be obtained:

$$P(\Phi_i\big||E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}|, |G_{\mathbf{h}}|, |G_{\mathbf{k}}|, |G_{\mathbf{l}}|) \simeq [2\pi I_0(Q_j)]^{-1} \exp[Q_j \cos\Phi_j]$$

for $j = 1, \ldots, 8$.

The analytical expressions of Q_j are too intricate and are not given here (the reader is referred to the original paper). We only say that Q_j may be positive or negative, so that reliable triplet phase estimates near 0 or near π are possible: the larger $|Q_j|$, the more reliable the phase estimate.

A useful interpretation of the formulae in terms of experimental parameters was suggested by Fortier *et al.* (1984): according to them, distributions do not depend, as in the case of the traditional three-phase invariants, on the total number of atoms per unit cell but rather on the scattering difference between the native protein and the derivative (that is, on the scattering of the heavy atoms in the derivative).

Hauptman's formulae were generalized by Giacovazzo *et al.* (1988): the new expressions were able to take into account the resolution effects on distribution parameters. The formulae are completely general and include as special cases native protein and heavy-atom isomorphous derivatives as well as X-ray and neutron diffraction data. Their complicated algebraic forms are easily reduced to a simple expression in the case of a native protein heavy-atom derivative: in particular, the reliability parameter for Φ_1 is

$$Q_1 = 2[\sigma_3/\sigma_2^{3/2}]_{\rm p}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}| + 2[\sigma_3/\sigma_2^{3/2}]_{\rm H}\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{l}}, \qquad (2.2.10.2)$$

where indices p and H warn that parameters have to be calculated over protein atoms and over heavy atoms, respectively, and

$$\Delta = (F_{\rm d} - F_{\rm p})/(\textstyle\sum f_j^2)_{\rm H}^{1/2}.$$

Δ is a pseudo-normalized difference (with respect to the heavy-atom structure) between moduli of structure factors.

Equation (2.2.10.2) may be compared with Karle's (1983) algebraic rule: if the sign of $\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{l}}$ is plus then the value of Φ_1 is estimated to be zero; if its sign is minus then the expected value of Φ_1 is close to π. In practice Karle's rule agrees with (2.2.10.2) only if the Cochran-type term in (2.2.10.2) may be neglected. Furthermore, (2.2.10.2) shows that large reliability values do not depend on the triple product of structure-factor differences, but on the triple product of pseudo-normalized differences.

A similar mathematical approach has been applied to estimate quartet invariants *via* isomorphous data. The result may be summarized as follows: a quartet is a phase relationship of order $N_{\rm H}^{-1}$ (Giacovazzo & Siliqi, 1996*a*,*b*; see also Kyriakidis *et al.*, 1996), with reliability factor equal to

$$G = \frac{2\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{l}}\Delta_{\mathbf{h}+\mathbf{k}+\mathbf{l}}}{Q_4 N_{\mathrm{H}}} \times \left\{1 + \left(\Delta^2_{\mathbf{h}+\mathbf{k}} - 1\right) + \left(\Delta^2_{\mathbf{h}+\mathbf{l}} - 1\right) + \left(\Delta^2_{\mathbf{k}+\mathbf{l}} - 1\right)\right\}, \quad (2.2.10.3)$$

where Q_4 is a suitable normalizing factor.

As previously stressed, equations (2.2.10.2) and (2.2.10.3) are valid if the lack of isomorphism and the errors in the measurements are assumed to be negligible. At first sight this approach seems more appealing than the traditional two-step procedures, however it did not prove to be competitive with them. The main reason is the absence in the Hauptman and Giacovazzo approaches of a probabilistic treatment of the errors: such a treatment, on the contrary, is basic for the traditional SIR–MIR techniques [see Blow & Crick (1959) and Terwilliger & Eisenberg (1987) for two related approaches].

The problem of the errors in the probabilistic scenario defined by the joint probability distribution functions approach has recently been overcome by Giacovazzo *et al.* (2001). In their probabilistic calculations the following assumptions were made:

$$|F_{\mathrm{d}j}| \exp(i\varphi_j) = |F_{\mathrm{p}}| \exp(i\varphi_{\mathrm{p}}) + F_{\mathrm{H}j} \exp(i\varphi_{\mathrm{H}j}) + |\mu_j| \exp(i\theta_j), \quad (2.2.10.4)$$

where j refers to the jth derivative. $|\mu_j| \exp(i\theta_j)$ is the error, which can include model as well as measurement errors.

A more realistic expression for the reliability factor G of triplet invariants is obtained by including the expression (2.2.10.4) in the probabilistic approach. Then the reliability parameter of the triplet invariants is transformed into (Giacovazzo *et al.*, 2001)

$$G = 2[\sigma_3/\sigma_2^{3/2}]_{\mathrm{p}} R_{\mathrm{p}1} R_{\mathrm{p}2} R_{\mathrm{p}3} + 2[\sigma_3/\sigma_2^{3/2}]_{\mathrm{H}} \frac{\Delta_1\Delta_2\Delta_3}{[1 + (\sigma^2_{\mu 1})_{\mathrm{H}}][1 + (\sigma^2_{\mu 2})_{\mathrm{H}}][1 + (\sigma^2_{\mu 3})_{\mathrm{H}}]}, \quad (2.2.10.5)$$

where $(\sigma^2_{\mu})_{\mathrm{H}} = |\mu|^2/(\Sigma f_j^2)_{\mathrm{H}}$.

Equation (2.2.10.5) suggests how the error influences the reliability of the triplet estimate: even quite a small value of $|\mu|^2$ may be critical if the scattering power of the heavy-atom substructure is a very small percentage of the derivative scattering power.

A one-step procedure has been implemented in a computer program (Giacovazzo *et al.*, 2002): it has been shown that the method is able to derive automatically, from the experimental data and without any user intervention, good quality (*i.e.* perfectly interpretable) electron-density maps.

2.2.10.5. SIR–MIR case: the two-step procedure. Finding the heavy-atom substructure by direct methods

The first trials for finding the heavy-atom substructure were based on the following assumption: the modulus of the isomorphous difference,

$$\Delta F = |F_{\mathrm{d}}| - |F_{\mathrm{p}}|,$$

is assumed at a first approximation as an estimate of the heavy-atom structure factor F_{H}. Perutz (1956) approximated $|F_{\mathrm{H}}|^2$ with the difference $(|F_{\mathrm{d}}|^2 - |F_{\mathrm{p}}|^2)$. Blow (1958) and Rossmann (1960) suggested a better approximation: $|F_{\mathrm{H}}|^2 \simeq |\Delta F|^2$. A deeper analysis was performed by Phillips (1966), Dodson & Vijayan (1971), Blessing & Smith (1999) and Grosse-Kunstleve & Brunger (1999). The use of direct methods requires the normalization of $|\Delta F|$ and application of the tangent formula (Wilson, 1978).

A sounder procedure has been suggested by Giacovazzo *et al.* (2004): they studied, for the SIR case, the joint probability distribution function

$$P(E_{\mathrm{H}}, E_{\mathrm{p}}, E_{\mathrm{d}})$$

under the following assumptions:

(*a*) the atomic positions of the native protein structure and the positions of the heavy atoms in the derivative structure are the primitive random variables of the probabilistic approach;

(*b*)

$$|F_{\mathrm{d}}| \exp(i\varphi_{\mathrm{d}}) = |F_{\mathrm{p}}| \exp(i\varphi_{\mathrm{p}}) + |F_{\mathrm{H}}| \exp(i\varphi_{\mathrm{H}}) + |\mu_{\mathrm{d}}| \exp(i\theta_{\mathrm{d}}) \quad (2.2.10.6)$$

is the structure factor of the derivative.

Then the conditional distribution $P(R_{\mathrm{H}}|R_{\mathrm{p}}, R_{\mathrm{d}})$ may be derived, from which $\langle R_{\mathrm{H}}|R_{\mathrm{p}}, R_{\mathrm{d}}\rangle$ may be obtained. In terms of structure factors

$$\langle |F_{\mathrm{H}}|^2 \rangle = \frac{\sum_{\mathrm{H}}}{\left(\sum_{\mathrm{H}} + \langle |\mu_{\mathrm{d}}|^2 \rangle\right)} \left[\langle |\mu_{\mathrm{d}}|^2 \rangle + \frac{\sum_{\mathrm{H}}}{\left(\sum_{\mathrm{H}} + \langle |\mu_{\mathrm{d}}|^2 \rangle\right)} \Delta^2_{\mathrm{iso}} \right]. \quad (2.2.10.7)$$

The effect of the errors on the evaluation of the moduli $|F_{\mathrm{H}}|^2$ may be easily derived: if $\langle |\mu_{\mathrm{d}}|^2 \rangle = 0$, equation (2.2.10.7) confirms Blow and Rossmann's approximation $\langle |F_{\mathrm{H}}|^2 \rangle \simeq |\Delta F|^2$. If $\langle |\mu_{\mathrm{d}}|^2 \rangle \neq 0$ Blow and Rossmann's estimate should be affected by a systematic error, increasing with $\langle |\mu_{\mathrm{d}}|^2 \rangle$.

2.2.10.6. SIR–MIR case: protein phasing by direct methods

Let us suppose that the various heavy-atom substructures have been determined. They may be used as additional prior information for a more accurate estimate of the φ_{p} values. To this purpose the distributions

$$P(E_{\mathrm{p}}, \mathbf{E}'_{\mathrm{d}}|\mathbf{E}'_{\mathrm{H}}) \equiv P(E_{\mathrm{p}}, E'_{\mathrm{d}1}, \ldots, E'_{\mathrm{d}n}|E'_{\mathrm{H}1}, \ldots, E'_{\mathrm{H}n}) \quad (2.2.10.8)$$

may be used under the assumption (2.2.10.6). $E'_{\mathrm{d}j}$ and $E'_{\mathrm{H}j}$, for $j = 1, \ldots, n$, are the structure factors of the jth derivative and of the jth heavy-atom substructure, respectively, both normalized with respect to the protein. Any joint probability density (2.2.10.8) may be reliably approximated by a multidimensional Gaussian distribution (Giacovazzo & Siliqi, 2002), from which the following conditional distribution is obtained:

$$P(\varphi_{\mathrm{p}}|R_{\mathrm{p}}, \mathbf{R}'_{\mathrm{d}}, \mathbf{E}'_{\mathrm{H}}) \simeq [2\pi I_0(G)]^{-1} \exp[\alpha_{\mathrm{p}} \cos(\varphi_{\mathrm{p}} - \theta_{\mathrm{p}})]$$

where θ_{p}, the expected value of φ_{p}, is given by

$$\tan\theta_{\mathrm{p}} = \frac{\sum_{j=1}^{n} G_j \sin\varphi_{\mathrm{H}j}}{\sum_{j=1}^{n} G_j \cos\varphi_{\mathrm{H}j}} = \frac{T}{B}$$

and $G_j = 2|F_{\mathrm{H}j}|\Delta F/\mu_j^2$.

$\alpha_{\mathrm{p}} = (T^2 + B^2)^{1/2}$ is the reliability factor of the phase estimate. A robust phasing procedure has been established which, starting from the observed moduli $|F_{\mathrm{p}}|, |F_{\mathrm{d}j}|, j = 1, \ldots, n$, is able to automatically provide, without any user intervention, a high-quality electron-density map of the protein (Giacovazzo *et al.*, 2002).

2.2.10.7. Integration of anomalous-dispersion techniques with direct methods

If the frequency of the radiation is close to an absorption edge of an atom, then that atom will scatter the X-rays anomalously (see Chapter 2.4) according to $f = f' + if''$. This results in the breakdown of Friedel's law. It was soon realized that the Bijvoet difference could also be used in the determination of phases (Peerdeman & Bijvoet, 1956; Ramachandran & Raman, 1956;

Okaya & Pepinsky, 1956). Since then, a great deal of work has been done both from algebraic (see Chapter 2.4) and from probabilistic points of view. In this section we are only interested in the second.

SAD (single anomalous dispersion) and MAD (multiple anomalous dispersion) techniques can be used. Both are characterized by one protein structure and one anomalous-scatterer substructure. The experimental diffraction data differ only because of the different anomalous scattering (not because of different anomalous-scatterer substructures). In the MAD case the anomalous-scatterer substructure is in some way 'overdetermined' by the data and, therefore, it is more convenient to use a two-step procedure: first define the positions of the anomalous scatterers, and then estimate the protein phase values. For completeness, we describe the one-step procedures in Section 2.2.10.8. These are based on the estimation of the structure invariants and on the application of the tangent formula. The two-step procedures are described in the Sections 2.2.10.9 and 2.2.10.10.

2.2.10.8. The SAD case: the one-step procedures

Probability distributions of diffraction intensities and of selected functions of diffraction intensities for dispersive structures have been given by various authors [Parthasarathy & Srinivasan (1964), see also Srinivasan & Parthasarathy (1976) and relevant literature cited therein]. We describe here some probabilistic formulae for estimating invariants of low order.

(*a*) *Estimation of two-phase structure invariants.* The conditional probability distribution of $\Phi = \varphi_{\mathbf{h}} + \varphi_{-\mathbf{h}}$ given $R_{\mathbf{h}}$ and $G_{\mathbf{h}}$ (normalized moduli of $F_{\mathbf{h}}$ and $F_{-\mathbf{h}}$, respectively) (Hauptman, 1982*b*; Giacovazzo, 1983*b*) is

$$P(\Phi|R_{\mathbf{h}}, G_{\mathbf{h}}) \simeq [2\pi I_0(Q)]^{-1} \exp[Q\cos(\Phi - q)], \qquad (2.2.10.9)$$

where

$$\begin{aligned} Q &= \frac{2R_{\mathbf{h}}G_{\mathbf{h}}}{\sqrt{c}}[c_1^2 + c_2^2]^{1/2}, \\ \cos q &= \frac{c_1}{[c_1^2 + c_2^2]^{1/2}}, \qquad \sin q = \frac{c_2}{[c_1^2 + c_2^2]^{1/2}}, \\ c_1 &= \sum_{j=1}^{N}(f_j'^2 - f_j''^2)/\sum, \\ c_2 &= 2\sum_{j=1}^{N} f_j' f_j''/\sum, \\ c &= [1 - (c_1^2 + c_2^2)]^2, \\ \sum &= \sum_{j=1}^{N}(f_j'^2 + f_j''^2). \end{aligned}$$

q is the most probable value of Φ: a large value of the parameter Q suggests that the phase relation $\Phi = q$ is reliable. Large values of Q are often available in practice: q, however, may be considered an estimate of $|\Phi|$ rather than of Φ because the enantiomorph is not fixed in (2.2.10.9). A formula for the estimation of Φ in centrosymmetric structures has been provided by Giacovazzo (1987).

(*b*) *Estimation of triplet invariants.* Kroon *et al.* (1977) first incorporated anomalous diffraction in order to estimate triplet invariants. Their work was based on an analysis of the complex double Patterson function. Subsequent probabilistic considerations (Heinermann *et al.*, 1978) confirmed their results, which can be so expressed:

$$\sin\bar{\Phi} = \frac{|\tau|^2 - |\bar{\tau}|^2}{4\tau''[\frac{1}{2}(|\tau|^2 + |\bar{\tau}|^2) - |\tau''|^2]^{1/2}}, \qquad (2.2.10.10)$$

where $(\mathbf{h} + \mathbf{k} + \mathbf{l} = 0)$,

$$\begin{aligned} \tau &= E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}} = R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{l}}\exp(i\Phi_{\mathbf{h},\mathbf{k}}), \\ \bar{\tau} &= E_{-\mathbf{h}}E_{-\mathbf{k}}E_{-\mathbf{l}} = G_{\mathbf{h}}G_{\mathbf{k}}G_{\mathbf{l}}\exp(i\Phi_{\bar{\mathbf{h}},\bar{\mathbf{k}}}), \\ \bar{\Phi} &= \tfrac{1}{2}(\Phi_{\mathbf{h},\mathbf{k}} - \Phi_{\bar{\mathbf{h}},\bar{\mathbf{k}}}), \end{aligned}$$

and τ'' is the contribution of the imaginary part of τ, which may be approximated in favourable conditions by

$$\begin{aligned} \tau'' = 2f''[&f_{\mathbf{h}}'f_{\mathbf{k}}' + f_{\mathbf{h}}'f_{\mathbf{l}}' + f_{\mathbf{k}}f_{\mathbf{l}}] \\ &\times [1 + S(R_{\mathbf{h}}^2 + R_{\mathbf{k}}^2 + R_{\mathbf{l}}^2 - 3)], \end{aligned}$$

where S is a suitable scale factor.

Equation (2.2.10.10) gives two possible values for $\bar{\Phi}$ (Φ and $\pi - \Phi$). Only if $R_{\mathbf{h}}R_{\mathbf{k}}R_{\overline{\mathbf{h}+\mathbf{k}}}$ is large enough may this phase ambiguity be resolved by choosing the angle nearest to zero.

The evaluation of triplet phases by means of anomalous dispersion has been further pursued by Hauptman (1982*b*) and independently by Giacovazzo (1983*b*). Owing to the breakdown of Friedel's law there are eight distinct triplet invariants which can contemporaneously be exploited:

$$\begin{aligned} \Phi_1 &= \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}, & \Phi_2 &= -\varphi_{-\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} \\ \Phi_3 &= \varphi_{\mathbf{h}} - \varphi_{-\mathbf{k}} + \varphi_{\mathbf{l}}, & \Phi_4 &= \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{-\mathbf{l}} \\ \Phi_5 &= \varphi_{-\mathbf{h}} + \varphi_{-\mathbf{k}} + \varphi_{-\mathbf{l}}, & \Phi_6 &= -\varphi_{\mathbf{h}} + \varphi_{-\mathbf{k}} + \varphi_{-\mathbf{l}} \\ \Phi_7 &= \varphi_{-\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{-\mathbf{l}}, & \Phi_8 &= \varphi_{-\mathbf{h}} + \varphi_{-\mathbf{k}} - \varphi_{\mathbf{l}}. \end{aligned}$$

Given

$$\begin{aligned} R_1 &= |E_{\mathbf{h}}| & R_2 &= |E_{\mathbf{k}}| & R_3 &= |E_{\mathbf{h}+\mathbf{k}}| \\ G_1 &= |E_{-\mathbf{h}}| & G_2 &= |E_{-\mathbf{k}}| & R_3 &= |E_{-\mathbf{h}-\mathbf{k}}| \\ \varphi_1 &= \varphi_{\mathbf{h}} & \varphi_2 &= \varphi_{\mathbf{k}} & \varphi_3 &= \varphi_{\mathbf{h}+\mathbf{k}} \\ \psi_1 &= \varphi_{-\mathbf{h}} & \psi_2 &= \varphi_{-\mathbf{k}} & \psi_3 &= \varphi_{-\mathbf{h}-\mathbf{k}}, \end{aligned}$$

Hauptman and Giacovazzo found the following conditional distribution:

$$P(\Phi|R_j, G_j, j = 1, 2, 3) \simeq \left[2\pi I_0(\Omega)\right]^{-1}\exp[\Omega\cos(\Phi - \omega)]. \qquad (2.2.10.11)$$

The definitions of Ω and ω are rather extensive and so the reader is referred to the published papers. We only add that Ω is always positive and that ω, the expected value of Φ, may lie anywhere between 0 and 2π. Understanding the role of the various parameters in equation (2.2.10.11) is not easy. Giacovazzo *et al.* (2003) found an equivalent simpler expression from which interpretable estimates of the parameters were obtained. In the same paper the limitations of the approach (*versus* the two-step procedures) were clarified.

2.2.10.9. SAD–MAD case: the two-step procedures. Finding the anomalous-scatterer substructure by direct methods

The anomalous-scatterer substructure is traditionally determined by the techniques suggested by Karle and Hendrickson (Karle, 1980*b*; Hendrickson, 1985; Pähler *et al.*, 1990; Terwilliger, 1994). The introduction of selenium into proteins as selenomethionine encouraged the second-generation direct methods programs [*Shake and Bake* by Miller *et al.* (1994); *Half bake* by Sheldrick (1998); *SIR2000-N* by Burla *et al.* (2001); *ACORN* by Foadi *et al.* (2000)] to locate Se atoms. Since the number of Se atoms may be quite large (up to 200), direct methods rather than Patterson techniques seem to be preferable. *Shake and Bake*, *Half Bake* and *ACORN* obtain the coordinates of the anomalous scatterers from a single-wavelength set of data. When more sets of diffraction data are available the solutions obtained by the other sets are used to confirm the correct solution.

A different approach has been suggested in two recent papers (Burla *et al.*, 2002; Burla, Carrozzini *et al.*, 2003): the estimates of the amplitudes of the structure factors of the anomalously scattering substructure are derived, *via* the rigorous method of the joint probability distribution functions, from the experimental diffraction moduli relative to n wavelengths. To do that, first the joint distribution

$$\begin{aligned} P_n = P(&A_{\rm oa}, A_1^+, A_2^+, \ldots, A_n^+, A_1^-, A_2^-, \ldots, A_n^-, \\ &B_{\rm oa}, B_1^+, B_2^+, \ldots, B_n^+, B_1^-, B_2^-, \ldots, B_n^-) \\ &= \pi^{-(2n+1)}(\det\, K)^{1/2} \exp(-\tfrac{1}{2}\mathbf{T}^T\mathbf{K}^{-1}\mathbf{T}) \end{aligned}$$

is calculated, where $A_{\rm oa}, B_{\rm oa}, E_{\rm oa}, A_i^+, B_i^+, A_i^-, B_i^-$ are the real and imaginary components of $E_{\rm oa}$, E_i^+, E_i^-, respectively, $\mathbf{K}$ is a symmetric square matrix of order $(4n + 2)$, $\mathbf{K}^{-1} = \{\lambda_{ij}\}$ is its inverse, and $\mathbf{T}$ is a suitable vector with components defined in terms of the variables $A_{\rm oa}, A_1^+, A_2^+, \ldots, B_n^-$. $E_{\rm oa}$ is the normalized structure factor of the anomalous scatterer substructure calculated by neglecting anomalous scattering components. Then the conditional distribution

$$P(R_{\rm oa}|R_1, \ldots, R_n, G_1, \ldots, G_n)$$

is derived, from which

$$\langle R_{\rm oa}|R_1, \ldots, G_n\rangle = \tfrac{1}{2}(\pi/\lambda_{11})^{1/2}[1 + 4X^2/(\pi\lambda_{11})]^{1/2} \quad (2.2.10.12)$$

is obtained, where

$$\begin{aligned} X^2 &= Q_1^2 + Q_2^2 \\ Q_1 &= \lambda_{12}R_1 + \lambda_{13}R_2 + \ldots + \lambda_{1,n+1}R_n + \lambda_{1,n+2}G_1 + \ldots \\ &\quad + \lambda_{1,2n+1}G_n \\ Q_2 &= \lambda_{1,2n+3}R_1 + \lambda_{1,2n+4}R_2 + \ldots + \lambda_{1,3n+2}R_n + \ldots - \lambda_{1,3n+3}G_1 \\ &\quad - \ldots - \lambda_{1,4n+2}G_n. \end{aligned}$$

The standard deviation of the estimate is also calculated:

$$\sigma_{R_{\rm oa}} = \left[\langle R_{\rm oa}^2|\ldots\rangle - \langle R_{\rm oa}|\ldots\rangle^2\right]^{1/2} = \left[\left(1 - \frac{\pi}{4}\right)\lambda_{11}^{-1}\right]^{1/2},$$

from which

$$\frac{\langle R_{\rm oa}|\ldots\rangle}{\sigma_{R_{\rm oa}}} = \left[\frac{(\pi/4) + (X^2)/\lambda_{11}}{1 - (\pi/4)}\right]^{1/2}. \quad (2.2.10.13)$$

The advantage of the above approach is that the estimates can simultaneously exploit both the anomalous and the dispersive differences. The computing procedure proposed by Burla, Carrozzini *et al.* (2003) is the following:

(i) The sets S_j, $j = 1, \ldots, n$, of the observed magnitudes (say $|F^+|$, $|F^-|$) are stored for all the n wavelengths.

(ii) The Wilson method is applied to put the sets S_j on their absolute scales.

(iii) Equations (2.2.10.12) and (2.2.10.13) are applied to obtain the values $\langle R_{\rm oa}|\ldots\rangle$ and $\langle R_{\rm oa}|\ldots\rangle/\sigma_{R_{\rm oa}}$.

(iv) The triplet invariants involving the reflections with the highest $\langle R_{\rm oa}|\ldots\rangle/\sigma_{R_{\rm oa}}$ values are evaluated and the tangent formula is applied *via* a random starting approach.

(v) The direct-space refinement techniques of *SIR2002* (Burla, Camalli *et al.*, 2003) are used to extend the phase information to a larger set of reflections: only 30% of the reflections with the smallest values of $\langle R_{\rm oa}|\ldots\rangle$ remain unphased. Automatic cycles of least-squares refinement improve the substructure model provided by the trial solutions.

(vi) Suitable figures of merit are used to recognize the correct substructure models.

The application of the above procedure to several MAD cases showed that the various wavelength combinations are not equally informative. A criterion based on the correlation among the various $\Delta_{\rm ano}$ values was also provided (see also Schneider & Sheldrick, 2002) for predicting the most informative combinations.

2.2.10.10. SAD–MAD case: protein phasing by direct methods

Once the anomalous-scatterer substructure has been found, the corresponding structure factors $E_{a1}^+, \ldots, E_{an}^+, E_{a1}^-, \ldots, E_{an}^-$ are known in modulus and phase. Then the conditional joint probability distribution

$$P\left(E_1^+, \ldots, E_n^+, E_1^-, \ldots, E_n^-|E_{a1}^+, \ldots, E_{an}^+, E_{a1}^-, \ldots, E_{an}^-\right)$$

may be calculated (Giacovazzo & Siliqi, 2004), from which the conditional distribution

$$P\left(\varphi_1^+|E_{ai}^+, E_{ai}^-, R_i, G_i, i = 1, \ldots, 2\right)$$

may be derived.

It has been shown that the most probable phase of φ_1^+, say θ_1^+, is the phase of the vector

$$\begin{aligned} &\sum_{j=1}^{n}\left[w_j^+E_{aj}^+ + w_j^-E_{aj}^{-*}\right] \\ &\quad + \sum_{j,p=1,p>j}^{n}\left[w_{jp}(E_{aj}^+ - E_{ap}^+) + w_{n+j,n+p}\left(E_{aj}^{-*} - E_{ap}^{-*}\right)\right] \\ &\quad + \sum_{j,p=1}^{n} w_{j,n+p}\left(E_{aj}^+ - E_{ap}^{-*}\right)] \end{aligned} \quad (2.2.10.14)$$

and the reliability parameter of the phase estimate is nothing other than the modulus of (2.2.10.14). The first term in (2.2.10.14) is a Sim-like contribution; the other terms, through the weights w, take into account the errors and the experimental differences $(R_j - R_p)$, $(G_j - G_p)$ and $(R_j - G_p)$.

References

Allegra, G. (1979). *Derivation of three-phase invariants from the Patterson function. Acta Cryst.* A**35**, 213–220.

Altomare, A., Burla, M. C., Camalli, M., Cascarano, G. L., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Spagna, R. (1999). *SIR97: a new tool for crystal structure determination and refinement. J. Appl. Cryst.* **32**, 115–119.

Anzenhofer, K. & Hoppe, W. (1962). *Phys. Verh. Mosbach.* **13**, 119.

Ardito, G., Cascarano, G., Giacovazzo, C. & Luić, M. (1985). *1-Phase seminvariants and Harker sections. Z. Kristallogr.* **172**, 25–34.

Argos, P. & Rossmann, M. G. (1980). *Molecular replacement method.* In *Theory and Practice of Direct Methods in Crystallography,* edited by M. F. C. Ladd & R. A. Palmer, pp. 381–389. New York: Plenum.

Avrami, M. (1938). *Direct determination of crystal structure from X-ray data. Phys. Rev.* **54**, 300–303.

Baggio, R., Woolfson, M. M., Declercq, J.-P. & Germain, G. (1978). *On the application of phase relationships to complex structures. XVI. A random approach to structure determination. Acta Cryst.* A**34**, 883–892.

Banerjee, K. (1933). *Determination of the signs of the Fourier terms in complete crystal structure analysis. Proc. R. Soc. London Ser. A*, **141**, 188–193.

Bertaut, E. F. (1955*a*). *La méthode statistique en cristallographie. I. Acta Cryst.* **8**, 537–543.

Bertaut, E. F. (1955*b*). *La méthode statistique en cristallographie. II. Quelques applications. Acta Cryst.* **8**, 544–548.

Bertaut, E. F. (1960). *Ordre logarithmique des densités de répartition. I. Acta Cryst.* **13**, 546–552.

Beurskens, P. T., Beurskens, G., de Gelder, R., Garcia-Granda, S., Gould, R. O., Israel, R. & Smits, J. M. M. (1999). *The DIRDIF-99 program system.* Crystallography Laboratory, University of Nijmegen, The Netherlands.

Beurskens, P. T., Gould, R. O., Bruins Slot, H. J. & Bosman, W. P. (1987). *Translation functions for the positioning of a well oriented molecular fragment. Z. Kristallogr.* **179**, 127–159.

Beurskens, P. T., Prick, A. J., Doesburg, H. M. & Gould, R. O. (1979). *Statistical properties of normalized difference-structure factors for non-centrosymmetric structures. Acta Cryst.* A**35**, 765–772.

Bhuiya, A. K. & Stanley, E. (1963). *The refinement of atomic parameters by direct calculation of the minimum residual. Acta Cryst.* **16**, 981–984.

Blessing, R. H. & Smith, G. D. (1999). *Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. J. Appl. Cryst.* **32**, 664–670.

Blow, D. M. (1958). *An X-ray examination of some crystal forms of pig and rabbit haemoglobin. Acta Cryst.* **11**, 125–126.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Böhme, R. (1982). *Direkte Methoden für Strukturen mit Uberstruktur-effekten. Acta Cryst.* A**38**, 318–326.

Bouman, J. (1956). *A general theory of inequalities. Acta Cryst.* **9**, 777–780.

Bricogne, G. (1984). *Maximum entropy and the foundation of direct methods. Acta Cryst.* A**40**, 410–415.

Britten, P. L. & Collins, D. M. (1982). *Information theory as a basis for the maximum determinant. Acta Cryst.* A**38**, 129–132.

Buerger, M. J. (1959). *Vector Space and its Applications in Crystal Structure Investigation.* New York: John Wiley.

Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. (2005). *SIR2004: an improved tool for crystal structure determination and refinement. J. Appl. Cryst.* **38**, 381–388.

Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Polidori, G. (2004). *Ab initio protein phasing: the Patterson deconvolution method in SIR2002. J. Appl. Cryst.* **37**, 258–264.

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2001). *SIR2000-N, a program for large and small crystal structures. J. Appl. Cryst.* **34**, 523–526.

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2003). *SIR2002: the program. J. Appl. Cryst.* **36**, 1103.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2003). *SAD or MAD phasing: location of the anomalous scatterers. Acta Cryst.* D**59**, 662–669.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Siliqi, D. (2002). *MAD phasing: probabilistic estimate of* $|F_{oa}|$. *Acta Cryst.* D**58**, 928–935.

Burla, M. C., Cascarano, G., Giacovazzo, C., Nunzi, A. & Polidori, G. (1987). *A weighting scheme for tangent formula development. III. The weighting scheme of the SIR program. Acta Cryst.* A**43**, 370–374.

Busetta, B. (1976). *An example of the use of quartet and triplet structure invariants when enantiomorph discrimination is difficult. Acta Cryst.* A**32**, 139–143.

Busetta, B., Giacovazzo, C., Burla, M. C., Nunzi, A., Polidori, G. & Viterbo, D. (1980). *The SIR program. I. Use of negative quartets. Acta Cryst.* A**36**, 68–74.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005*a*). *Phasing at resolution higher than the experimental resolution. Acta Cryst.* D**61**, 556–565.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005*b*). *Ab initio phasing at resolution higher than experimental resolution. Acta Cryst.* D**61**, 1080–1087.

Camalli, M., Giacovazzo, C. & Spagna, R. (1985). *From a partial to the complete crystal structure. II. The procedure and its applications. Acta Cryst.* A**41**, 605–613.

Cascarano, G. & Giacovazzo, C. (1983). *One-phase seminvariants of first rank. I. Algebraic considerations. Z. Kristallogr.* **165**, 169–174.

Cascarano, G., Giacovazzo, C., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *The distribution of* $\alpha_{\mathbf{h}}$. *Acta Cryst.* A**40**, 389–394.

Cascarano, G., Giacovazzo, C., Calabrese, G., Burla, M. C., Nunzi, A., Polidori, G. & Viterbo, D. (1984). *One-phase seminvariants of first rank. II. Probabilistic considerations. Z. Kristallogr.* **167**, 37–47.

Cascarano, G., Giacovazzo, C., Camalli, M., Spagna, R., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *The method of representations of structure seminvariants. The strengthening of triplet relationships. Acta Cryst.* A**40**, 278–283.

Cascarano, G., Giacovazzo, C. & Luić, M. (1985*a*). *Non-crystallographic translational symmetry: effects on diffraction-intensity statistics.* In *Structure and Statistics in Crystallography*, edited by A. J. C. Wilson, pp. 67–77. Guilderland, USA: Adenine Press.

Cascarano, G., Giacovazzo, C. & Luić, M. (1985*b*). *Direct methods and superstructures. I. Effects of the pseudotranslation on the reciprocal space. Acta Cryst.* A**41**, 544–551.

Cascarano, G., Giacovazzo, C. & Luić, M. (1987). *Direct methods and structures showing superstructure effects. II. A probabilistic theory of triplet invariants. Acta Cryst.* A**43**, 14–22.

Cascarano, G., Giacovazzo, C. & Luić, M. (1988*a*). *Direct methods and structures showing superstructure effects. III. A general mathematical model. Acta Cryst.* A**44**, 176–183.

Cascarano, G., Giacovazzo, C. & Luić, M. (1988*b*). *Direct methods and structures showing superstructure effects. IV. A new approach for phase solution. Acta Cryst.* A**44**, 183–188.

Cascarano, G., Giacovazzo, C., Luić, M., Pifferi, A. & Spagna, R. (1987). *1-Phase seminvariants and Harker sections. II. A new procedure. Z. Kristallogr.* **179**, 113–125.

Cascarano, G., Giacovazzo, C. & Viterbo, D. (1987). *Figures of merit in direct methods: a new point of view. Acta Cryst.* A**43**, 22–29.

Castellano, E. E., Podjarny, A. D. & Navaza, J. (1973). *A multivariate joint probability distribution of phase determination. Acta Cryst.* A**29**, 609–615.

Cochran, W. (1955). *Relations between the phases of structure factors. Acta Cryst.* **8**, 473–478.

Cochran, W. & Douglas, A. S. (1957). *The use of a high-speed digital computer for the direct determination of crystal structure. II. Proc. R. Soc. London Ser. A*, **243**, 281–288.

Cochran, W. & Woolfson, M. M. (1955). *The theory of sign relations between structure factors. Acta Cryst.* **8**, 1–12.

Coulter, C. L. & Dewar, R. B. K. (1971). *Tangent formula applications in protein crystallography: an evaluation. Acta Cryst.* B**27**, 1730–1740.

Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.

Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Isaacs, N. W., Sakabe, K. & Sakabe, N. (1975). *The high resolution structure of insulin: a comparison of results obtained from least-squares phase refinement and difference Fourier refinement. Acta Cryst.* A**31**, S21.

Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *On the application of phase relationships to complex structures. XXIV. The Sayre tangent formula. Acta Cryst.* A**41**, 286–290.

Debaerdemaeker, T. & Woolfson, M. M. (1972). *On the application of phase relationships to complex structures. IV. The coincidence method applied to general phases. Acta Cryst.* A**28**, 477–481.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *On the application of phase relationships to complex structures. XXII. Techniques for random refinement. Acta Cryst.* A**39**, 193–196.

Declercq, J.-P., Germain, G., Main, P. & Woolfson, M. M. (1973). *On the application of phase relationships to complex structures. V. Finding the solution. Acta Cryst.* A**29**, 231–234.

Declercq, J.-P., Germain, G. & Woolfson, M. M. (1975). *On the application of phase relationships to complex structures. VIII. Extension of the magic-integer approach. Acta Cryst.* A**31**, 367–372.

De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Use of the negative quartet cosine invariants as a phasing figure of merit: NQEST. Acta Cryst.* A**31**, 472–479.

DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Structure solution by minimal-function phase refinement and Fourier filtering. I. Theoretical basis. Acta Cryst.* A**50**, 203–210.

Dodson, E. & Vijayan, M. (1971). *The determination and refinement of heavy-atom parameters in protein heavy-atom derivatives. Some model calculations using acentric reflexions. Acta Cryst.* B**27**, 2402–2411.

Egert, E. (1983). *Patterson search – an alternative to direct methods. Acta Cryst.* A**39**, 936–940.

Egert, E. & Sheldrick, G. M. (1985). *Search for a fragment of known geometry by integrated Patterson and direct methods. Acta Cryst.* A**41**, 262–268.

Eller, G. von (1973). *Génération de formules statistiques entre facteurs de structure: la méthode du polynome. Acta Cryst*. A**29**, 63–67.

Fan, H.-F. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

Fan, H. F., Hao, Q. & Woolfson, M. M. (1991). *Proteins and direct methods. Z. Kristallogr.* **197**, 197–208.

Fan, H.-F., Yao, J.-X., Main, P. & Woolfson, M. M. (1983). *On the application of phase relationships to complex structures. XXIII. Automatic determination of crystal structures having pseudo-translational symmetry by a modified MULTAN procedure. Acta Cryst.* A**39**, 566–569.

Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *A flexible and efficient procedure for the solution and phase refinement of protein structures. Acta Cryst.* D**56**, 1137–1147.

Fortier, S. & Hauptman, H. (1977). *Quintets in $P\bar{1}$: probabilistic theory of the five-phase structure invariant in the space group $P\bar{1}$. Acta Cryst.* A**33**, 829–833.

Fortier, S., Weeks, C. M. & Hauptman, H. (1984). *On integrating the techniques of direct methods and isomorphous replacement. III. The three-phase invariant for the native and two-derivative case. Acta Cryst.* A**40**, 646–651.

Freer, A. A. & Gilmore, C. J. (1980). *The use of higher invariants in MULTAN. Acta Cryst.* A**36**, 470–475.

French, S. & Wilson, K. (1978). *On the treatment of negative intensity observations. Acta Cryst.* A**34**, 517–525.

Gelder, R. de (1992). Thesis. University of Leiden, The Netherlands.

Gelder, R. de, de Graaff, R. A. G. & Schenk, H. (1990). *On the construction of Karle–Hauptman matrices. Acta Cryst.* A**46**, 688–692.

Gelder, R. de, de Graaff, R. A. G. & Schenk, H. (1993). *Automatic determination of crystal structures using Karle–Hauptman matrices. Acta Cryst.* A**49**, 287–293.

Germain, G., Main, P. & Woolfson, M. M. (1970). *On the application of phase relationships to complex structures. II. Getting a good start. Acta Cryst.* B**26**, 274–285.

Germain, G., Main, P. & Woolfson, M. M. (1971). *The application of phase relationships to complex structures. III. The optimum use of phase relationships. Acta Cryst.* A**27**, 368–376.

Giacovazzo, C. (1974). *A new scheme for seminvariant tables in all space groups. Acta Cryst.* A**30**, 390–395.

Giacovazzo, C. (1975). *A probabilistic theory in $P\bar{1}$ of the invariant $E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}E_{\mathbf{h+k+l}}$. Acta Cryst.* A**31**, 252–259.

Giacovazzo, C. (1976). *A probabilistic theory of the cosine invariant* $\cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} - \varphi_{\mathbf{h+k+l}})$. *Acta Cryst.* A**32**, 91–99.

Giacovazzo, C. (1977*a*). *A general approach to phase relationships: the method of representations. Acta Cryst.* A**33**, 933–944.

Giacovazzo, C. (1977*b*). *Strengthening of the triplet relationships. II. A new probabilistic approach in $P\bar{1}$. Acta Cryst.* A**33**, 527–531.

Giacovazzo, C. (1977*c*). *On different probabilistic approaches to quartet theory. Acta Cryst.* A**33**, 50–54.

Giacovazzo, C. (1977*d*). *Quintets in $P\bar{1}$ and related phase relationships: a probabilistic approach. Acta Cryst.* A**33**, 944–948.

Giacovazzo, C. (1977*e*). *A probabilistic theory of the coincidence method. I. Centrosymmetric space groups. Acta Cryst.* A**33**, 531–538.

Giacovazzo, C. (1977*f*). *A probabilistic theory of the coincidence method. II. Non-centrosymmetric space groups. Acta Cryst.* A**33**, 539–547.

Giacovazzo, C. (1978). *The estimation of the one-phase structure seminvariants of first rank by means of their first and second representation. Acta Cryst.* A**34**, 562–574.

Giacovazzo, C. (1979*a*). *A probabilistic theory of two-phase seminvariants of first rank via the method of representations. III. Acta Cryst.* A**35**, 296–305.

Giacovazzo, C. (1979*b*). *A theoretical weighting scheme for tangent-formula development and refinement and Fourier synthesis. Acta Cryst.* A**35**, 757–764.

Giacovazzo, C. (1980*a*). *Direct Methods in Crystallography*. London: Academic Press.

Giacovazzo, C. (1980*b*). *The method of representations of structure seminvariants. II. New theoretical and practical aspects. Acta Cryst.* A**36**, 362–372.

Giacovazzo, C. (1980*c*). *Triplet and quartet relations: their use in direct procedures. Acta Cryst.* A**36**, 74–82.

Giacovazzo, C. (1983*a*). *From a partial to the complete crystal structure. Acta Cryst.* A**39**, 685–692.

Giacovazzo, C. (1983*b*). *The estimation of two-phase invariants in $P\bar{1}$ when anomalous scatterers are present. Acta Cryst.* A**39**, 585–592.

Giacovazzo, C. (1987). *One wavelength technique: estimation of centrosymmetrical two-phase invariants in dispersive structures. Acta Cryst.* A**43**, 73–75.

Giacovazzo, C. (1988). *New probabilistic formulas for finding the positions of correctly oriented atomic groups. Acta Cryst.* A**44**, 294–300.

Giacovazzo, C. (1998). *Direct Phasing in Crystallography*. New York: IUCr, Oxford University Press.

Giacovazzo, C., Cascarano, G. & Zheng, C.-D. (1988). *On integrating the techniques of direct methods and isomorphous replacement. A new probabilistic formula for triplet invariants. Acta Cryst.* A**44**, 45–51.

Giacovazzo, C., Guagliardi, A., Ravelli, R. & Siliqi, D. (1994). *Ab initio direct phasing of proteins: the limits. Z. Kristallogr.* **209**, 136–142.

Giacovazzo, C., Ladisa, M. & Siliqi, D. (2002). *Crystal structure solution of proteins by direct methods: an automatic procedure for SIR–MIR and SIRAS–MIRAS cases. Acta Cryst.* A**58**, 598–604.

Giacovazzo, C., Ladisa, M. & Siliqi, D. (2003). *The estimation of three-phase invariants when anomalous scatterers are present: the limits. Acta Cryst.* A**59**, 569–576.

Giacovazzo, C., Moustiakimov, M., Siliqi, D. & Pifferi, A. (2004). *Locating heavy atoms by integrating direct methods and SIR techniques. Acta Cryst.* A**60**, 233–238.

Giacovazzo, C. & Siliqi, D. (1996*a*). *On integrating direct methods and isomorphous replacement techniques. I. A distribution function for quartet invariants. Acta Cryst.* A**52**, 133–142.

Giacovazzo, C. & Siliqi, D. (1996*b*). *On integrating direct methods and isomorphous replacement techniques. II. The quartet invariant estimate. Acta Cryst.* A**52**, 143–151.

Giacovazzo, C. & Siliqi, D. (2002). *The method of joint probability distribution functions applied to SIR–MIR and to SIRAS–MIRAS cases. Acta Cryst.* A**58**, 590–597.

Giacovazzo, C. & Siliqi, D. (2004). *Phasing via SAD/MAD data: the method of the joint probability distribution functions. Acta Cryst.* D**60**, 73–82.

Giacovazzo, C., Siliqi, D. & García-Rodríguez, L. (2001). *On integrating direct methods and isomorphous-replacement techniques: triplet estimation and treatment of errors. Acta Cryst.* A**57**, 571–575.

Gillis, J. (1948). *Structure factor relations and phase determination. Acta Cryst.* **1**, 76–80.

Gilmore, C. J. (1984). *MITHRIL. An integrated direct-methods computer program. J. Appl. Cryst.* **17**, 42–46.

Goedkoop, J. A. (1950). *Remarks on the theory of phase limiting inequalities and equalities. Acta Cryst.* **3**, 374–378.

Gramlich, V. (1984). *The influence of rational dependence on the probability distribution of structure factors. Acta Cryst.* A**40**, 610–616.

Grant, D. F., Howells, R. G. & Rogers, D. (1957). *A method for the systematic application of sign relations. Acta Cryst.* **10**, 489–497.

Grosse-Kunstleve, R. W. & Brunger, A. T. (1999). *A highly automated heavy-atom search procedure for macromolecular structures. Acta Cryst.* D**55**, 1568–1577.

Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (1999). *Xtal3.6 crystallographic software.* http://xtal.sourceforge.net/.

Harker, D. & Kasper, J. S. (1948). *Phases of Fourier coefficients directly from crystal diffraction data. Acta Cryst.* **1**, 70–75.

Hauptman, H. (1964). *The role of molecular structure in the direct determination of phase. Acta Cryst.* **17**, 1421–1433.

Hauptman, H. (1965). *The average value of* $\exp\{2\pi i(\mathbf{h}\cdot\mathbf{r} + \mathbf{h}'\cdot\mathbf{r}')\}$. *Z. Kristallogr.* **121**, 1–8.

Hauptman, H. (1970). Communication at New Orleans Meeting of Am. Crystallogr. Assoc.

Hauptman, H. (1974). *On the identity and estimation of those cosine invariants*, $\cos(\varphi_{\mathbf{m}} + \varphi_{\mathbf{n}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}})$, *which are probably negative. Acta Cryst.* A**30**, 472–476.

Hauptman, H. (1975). *A new method in the probabilistic theory of the structure invariants. Acta Cryst.* A**31**, 680–687.

Hauptman, H. (1982*a*). *On integrating the techniques of direct methods and isomorphous replacement. I. The theoretical basis. Acta Cryst.* A**38**, 289–294.

Hauptman, H. (1982*b*). *On integrating the techniques of direct methods with anomalous dispersion. I. The theoretical basis. Acta Cryst.* A**38**, 632–641.

Hauptman, H., Fisher, J., Hancock, H. & Norton, D. A. (1969). *Phase determination for the estriol structure. Acta Cryst.* B**25**, 811–814.

Hauptman, H. & Green, E. A. (1976). *Conditional probability distributions of the four-phase structure invariant $\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{m}}$ in $P\bar{1}$. Acta Cryst.* A**32**, 45–49.

Hauptman, H. & Green, E. A. (1978). *Pairs in $P2_1$: probability distributions which lead to estimates of the two-phase structure seminvariants in the vicinity of $\pi/2$. Acta Cryst.* A**34**, 224–229.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* Am. Crystallogr. Assoc. Monograph No. 3. Dayton, Ohio: Polycrystal Book Service.

Hauptman, H. & Karle, J. (1956). *Structure invariants and seminvariants for non-centrosymmetric space groups. Acta Cryst.* **9**, 45–55.

Hauptman, H. & Karle, J. (1958). *Phase determination from new joint probability distributions: space group $P\bar{1}$. Acta Cryst.* **11**, 149–157.

Hauptman, H. & Karle, J. (1959). *Table 2. Equivalence classes, seminvariant vectors and seminvariant moduli for the centered centrosymmetric space groups, referred to a primitive unit cell. Acta Cryst.* **12**, 93–97.

Hauptman, H. A. (1991). In *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry. IUCr/Oxford University Press.

Hauptman, H. A., Xu, H., Weeks, C. M. & Miller, R. (1999). *Exponential Shake-and-Bake: theoretical basis and applications. Acta Cryst.* A**55**, 891–900.

Heinermann, J. J. L. (1977*a*). *The use of structural information in the phase probability of a triple product. Acta Cryst.* A**33**, 100–106.

Heinermann, J. J. L. (1977*b*). Thesis. University of Utrecht.

Heinermann, J. J. L., Krabbendam, H. & Kroon, J. (1979). *The joint probability distribution of the structure factors in a Karle–Hauptman matrix. Acta Cryst.* A**35**, 101–105.

Heinermann, J. J. L., Krabbendam, H., Kroon, J. & Spek, A. L. (1978). *Direct phase determination of triple products from Bijvoet inequalities. II. A probabilistic approach. Acta Cryst.* A**34**, 447–450.

Hendrickson, W. A. (1985). *Analysis of protein structure from diffraction measurement at multiple wavelengths. Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

Hendrickson, W. A., Love, W. E. & Karle, J. (1973). *Crystal structure analysis of sea lamprey hemoglobin at 2 Å resolution. J. Mol. Biol.* **74**, 331–361.

Hoppe, W. (1963). *Phase determination and zero points in the Patterson function. Acta Cryst.* **16**, 1056–1057.

Hughes, E. W. (1953). *The signs of products of structure factors. Acta Cryst.* **6**, 871.

Hull, S. E. & Irwin, M. J. (1978). *On the application of phase relationships to complex structures. XIV. The additional use of statistical information in tangent-formula refinement. Acta Cryst.* A**34**, 863–870.

Hull, S. E., Viterbo, D., Woolfson, M. M. & Shao-Hui, Z. (1981). *On the application of phase relationships to complex structures. XIX. Magic-integer representation of a large set of phases: the MAGEX procedure. Acta Cryst.* A**37**, 566–572.

International Tables for Crystallography (2001). Vol. F, *Macromolecular Crystallography*, edited by M. G. Rossmann & E. Arnold. Dordrecht: Kluwer Academic Publishers.

Jaynes, E. T. (1957). *Information theory and statistical mechanics. Phys. Rev.* **106**, 620–630.

Karle, J. (1970*a*). *An alternative form for $B_{3,0}$, a phase determining formula. Acta Cryst.* B**26**, 1614–1617.

Karle, J. (1970*b*). *Partial structures and use of the tangent formula and translation functions.* In *Crystallographic Computing*, pp. 155–164. Copenhagen: Munksgaard.

Karle, J. (1972). *Translation functions and direct methods. Acta Cryst.* B**28**, 820–824.

Karle, J. (1979). *Triple phase invariants: formula for centric case from fourth-order determinantal joint probability distributions. Proc. Natl Acad. Sci. USA*, **76**, 2089–2093.

Karle, J. (1980*a*). *Triplet phase invariants: formula for acentric case from fourth-order determinantal joint probability distributions. Proc. Natl Acad. Sci. USA*, **77**, 5–9.

Karle, J. (1980*b*). *Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. Int. J. Quantum Chem. Quantum Biol. Symp.* **7**, 357–367.

Karle, J. (1983). *A simple rule for finding and distinguishing triplet phase invariants with values near 0 or π with isomorphous replacement data. Acta Cryst.* A**39**, 800–805.

Karle, J. & Hauptman, H. (1950). *The phases and magnitudes of the structure factors. Acta Cryst.* **3**, 181–187.

Karle, J. & Hauptman, H. (1956). *A theory of phase determination for the four types of non-centrosymmetric space groups* 1*P*222, 2*P*22, $3P_1$2, $3P_2$2. *Acta Cryst.* **9**, 635–651.

Karle, J. & Hauptman, H. (1958). *Phase determination from new joint probability distributions: space group P*1. *Acta Cryst.* **11**, 264–269.

Karle, J. & Hauptman, H. (1961). *Seminvariants for non-centrosymmetric space groups with conventional centered cells. Acta Cryst.* **14**, 217–223.

Karle, J. & Karle, I. L. (1966). *The symbolic addition procedure for phase determination for centrosymmetric and non-centrosymmetric crystals. Acta Cryst.* **21**, 849–859.

Klug, A. (1958). *Joint probability distributions of structure factors and the phase problem. Acta Cryst.* **11**, 515–543.

Koch, M. H. J. (1974). *On the application of phase relationships to complex structures. IV. Automatic interpretation of electron-density maps for organic structures. Acta Cryst.* A**30**, 67–70.

Krabbendam, H. & Kroon, J. (1971). *A relation between structure factor, triple products and a single Patterson vector, and its application to sign determination. Acta Cryst.* A**27**, 362–367.

Kroon, J., Spek, A. L. & Krabbendam, H. (1977). *Direct phase determination of triple products from Bijvoet inequalities. Acta Cryst.* A**33**, 382–385.

Kyriakidis, C. E., Peschar, R. & Schenk, H. (1996). *The estimation of four-phase structure invariants using the single difference of isomorphous structure factors. Acta Cryst.* A**52**, 77–87.

Lajzérowicz, J. & Lajzérowicz, J. (1966). *Loi de distribution des facteurs de structure pour un répartition non uniforme des atomes. Acta Cryst.* **21**, 8–12.

Langs, D. A. (1985). *Translation functions: the elimination of structure-dependent spurious maxima. Acta Cryst.* A**41**, 305–308.

Lessinger, L. & Wondratschek, H. (1975). *Seminvariants for space groups $I\bar{4}2m$ and $I\bar{4}d$. Acta Cryst.* A**31**, 521.

Mackay, A. L. (1953). *A statistical treatment of superlattice reflexions. Acta Cryst.* **6**, 214–215.

Main, P. (1976). *Recent developments in the MULTAN system. The use of molecular structure.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 97–105. Copenhagen: Munksgaard.

Main, P. (1977). *On the application of phase relationships to complex structures. XI. A theory of magic integers. Acta Cryst.* A**33**, 750–757.

Main, P., Fiske, S. J., Germain, G., Hull, S. E., Declercq, J.-P., Lessinger, L. & Woolfson, M. M. (1999). *Crystallographic software: teXsan for Windows.* http://www.rigaku.com/downloads/journal/Vol15.1.1998/texsan.pdf.

Main, P. & Hull, S. E. (1978). *The recognition of molecular fragments in E maps and electron density maps. Acta Cryst.* A**34**, 353–361.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *SnB: crystal structure determination via shake-and-bake. J. Appl. Cryst.* **27**, 613–621.

Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *The use of MULTAN to locate the positions of anomalous scatterers. Acta Cryst.* A**45**, 715–718.

Narayan, R. & Nityananda, R. (1982). *The maximum determinant method and the maximum entropy method. Acta Cryst.* A**38**, 122–128.

Navaza, J. (1985). *On the maximum-entropy estimate of the electron density function. Acta Cryst.* A**41**, 232–244.

Navaza, J., Castellano, E. E. & Tsoucaris, G. (1983). *Constrained density modifications by variational techniques. Acta Cryst.* A**39**, 622–631.

Naya, S., Nitta, I. & Oda, T. (1964). *A study on the statistical method for determination of signs of structure factors. Acta Cryst.* **17**, 421–433.

Naya, S., Nitta, I. & Oda, T. (1965). *Affinement tridimensional du sulfanilamide β. Acta Cryst.* **19**, 734–747.

Nordman, C. E. (1985). *Introduction to Patterson search methods.* In *Crystallographic Computing 3. Data Collection, Structure Determination, Proteins and Databases*, edited by G. M. Sheldrick, G. Kruger & R. Goddard, pp. 232–244. Oxford: Clarendon Press.

Oda, T., Naya, S. & Taguchi, I. (1961). *Matrix theoretical derivation of inequalities. II. Acta Cryst.* **14**, 456–458.

Okaya, J. & Nitta, I. (1952). *Linear structure factor inequalities and the application to the structure determination of tetragonal ethylenediamine sulphate. Acta Cryst.* **5**, 564–570.

Okaya, Y. & Pepinsky, R. (1956). *New formulation and solution of the phase problem in X-ray analysis of non-centric crystals containing anomalous scatterers. Phys. Rev.* **103**, 1645–1647.

Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *New method in X-ray crystal structure determination involving the use of anomalous dispersion. Phys. Rev.* **98**, 1857–1858.

Oszlányi, G. & Süto, A. (2004). *Ab initio structure solution by charge flipping. Acta Cryst.* A**60**, 134–141.

Ott, H. (1927). *Zur Methodik der Struckturanalyse. Z. Kristallogr.* **66**, 136–153.

Pähler, A., Smith, J. L. & Hendrickson, W. A. (1990). *A probability representation for phase information from multiwavelength anomalous dispersion. Acta Cryst.* A**46**, 537–540.

Parthasarathy, S. & Srinivasan, R. (1964). *The probability distribution of Bijvoet differences. Acta Cryst.* **17**, 1400–1407.

Pavelcík, F. (1988). *Patterson-oriented automatic structure determination: getting a good start. Acta Cryst.* A**44**, 724–729.

Pavelcík, F., Kuchta, L. & Sivý, J. (1992). *Patterson-oriented automatic structure determination. Utilizing Patterson peaks. Acta Cryst.* A**48**, 791–796.

Peerdeman, A. F. & Bijvoet, J. M. (1956). *The indexing of reflexions in investigations involving the use of the anomalous scattering effect. Acta Cryst.* **9**, 1012–1015.

Perutz, M. F. (1956). *Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Cryst.* **9**, 867–873.

Phillips, D. C. (1966). *Advances in protein crystallography*. In *Advances in Structure Research by Diffraction Methods*, Vol. 2, edited by R. Brill & R. Mason, pp. 75–140. New York: John Wiley.

Piro, O. E. (1983). *Information theory and the phase problem in crystallography. Acta Cryst.* A**39**, 61–68.

Podjarny, A. D., Schevitz, R. W. & Sigler, P. B. (1981). *Phasing low-resolution macromolecular structure factors by matricial direct methods. Acta Cryst.* A**37**, 662–668.

Podjarny, A. D., Yonath, A. & Traub, W. (1976). *Application of multivariate distribution theory to phase extension for a crystalline protein. Acta Cryst.* A**32**, 281–292.

Rae, A. D. (1977). *The use of structure factors to find the origin of an oriented molecular fragment. Acta Cryst.* A**33**, 423–425.

Ramachandran, G. N. & Raman, S. (1956). *A new method for the structure analysis of non-centrosymmetric crystals. Curr. Sci. (India)*, **25**, 348.

Rango, C. de (1969). Thesis. Paris.

Rango, C. de, Mauguen, Y. & Tsoucaris, G. (1975). *Use of high-order probability laws in phase refinement and extension of protein structures. Acta Cryst.* A**31**, 227–233.

Rango, C. de, Mauguen, Y., Tsoucaris, G., Dodson, E. J., Dodson, G. G. & Taylor, D. J. (1985). *The extension and refinement of the 1.9 Å spacing isomorphous phases to 1.5 Å spacing in 2Zn insulin by determinantal methods. Acta Cryst.* A**41**, 3–17.

Rango, C. de, Tsoucaris, G. & Zelwer, C. (1974). *Phase determination from the Karle–Hauptman determinant. II. Connexion between inequalities and probabilities. Acta Cryst.* A**30**, 342–353.

Richardson, J. W. & Jacobson, R. A. (1987). *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 310–317. Oxford University Press.

Rogers, D., Stanley, E. & Wilson, A. J. C. (1955). *The probability distribution of intensities. VI. The influence of intensity errors on the statistical tests. Acta Cryst.* **8**, 383–393.

Rogers, D. & Wilson, A. J. C. (1953). *The probability distribution of X-ray intensities. V. A note on some hypersymmetric distributions. Acta Cryst.* **6**, 439–449.

Rossmann, M. G. (1960). *The accurate determination of the position and shape of heavy-atom replacement groups in proteins. Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961). *The position of anomalous scatterers in protein crystals. Acta Cryst.* **14**, 383–388.

Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). *The relative positions of independent molecules within the same asymmetric unit. Acta Cryst.* **17**, 338–342.

Sayre, D. (1952). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Sayre, D. (1953). *Double Patterson function. Acta Cryst.* **6**, 430–431.

Sayre, D. (1972). *On least-squares refinement of the phases of crystallographic structure factors. Acta Cryst.* A**28**, 210–212.

Sayre, D. & Toupin, R. (1975). *Major increase in speed of least-squares phase refinement. Acta Cryst.* A**31**, S20.

Schenk, H. (1973*a*). *Direct structure determination in P*1 *and other non-centrosymmetric symmorphic space groups. Acta Cryst.* A**29**, 480–481.

Schenk, H. (1973*b*). *The use of phase relationships between quartets of reflexions. Acta Cryst.* A**29**, 77–82.

Schneider, T. R. & Sheldrick, G. M. (2002). *Substructure solution with SHELXD. Acta Cryst.* D**58**, 1772–1779.

Sheldrick, G. M. (1990). *Phase annealing in SHELX-90: direct methods for larger structures. Acta Cryst.* A**46**, 467–473.

Sheldrick, G. M. (1992). *Crystallographic Computing 5*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 145–157. Oxford University Press.

Sheldrick, G. M. (1998). *SHELX: applications to macromolecules*. In *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.

Sheldrick, G. M. (2000). *The SHELX home page.* http://shelx.uni-ac.gwdg.de/SHELX/.

Sheldrick, G. M. & Gould, R. O. (1995). *Structure solution by iterative peaklist optimization and tangent expansion in space group P*1. *Acta Cryst.* B**51**, 423–431.

Sim, G. A. (1959). *The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atoms method for non-centrosymmetrical structures. Acta Cryst.* **12**, 813–815.

Simerska, M. (1956). *Czech. J. Phys.* **6**, 1.

Simonov, V. I. & Weissberg, A. M. (1970). *Calculation of the signs of structure amplitudes by a binary function section of interatomic vectors. Sov. Phys. Dokl.* **15**, 321–323.

Sint, L. & Schenk, H. (1975). *Phase extension and refinement in non-centrosymmetric structures containing large molecules. Acta Cryst.* A**31**, S22.

Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.

Taylor, D. J., Woolfson, M. M. & Main, P. (1978). *On the application of phase relationships to complex structures. XV. Magic determinants. Acta Cryst.* A**34**, 870–883.

Terwilliger, T. C. (1994). *MAD phasing: Bayesian estimates of* F_A. *Acta Cryst.* D**50**, 11–16.

Terwilliger, T. C. & Eisenberg, D. (1987). *Isomorphous replecement: effect of errors on the phase probability distribution. Acta Cryst.* A**43**, 6–13.

Tsoucaris, G. (1970). *A new method for phase determination. The maximum determinant rule. Acta Cryst.* A**26**, 492–499.

Van der Putten, N. & Schenk, H. (1977). *On the conditional probability of quintets. Acta Cryst.* A**33**, 856–858.

Vaughan, P. A. (1958). *A phase-determining procedure related to the vector-coincidence method. Acta Cryst.* **11**, 111–115.

Vermin, W. J. & de Graaff, R. A. G. (1978). *The use of Karle–Hauptman determinants in small-structure determinations. Acta Cryst.* A**34**, 892–894.

Vicković, I. & Viterbo, D. (1979). *A simple statistical treatment of unobserved reflexions. Application to two organic substances. Acta Cryst.* A**35**, 500–501.

Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Structure solution by minimal-function phase refinement and Fourier filtering. II. Implementation and applications. Acta Cryst.* A**50**, 210–220.

Weeks, C. M. & Miller, R. (1999). *The design and implementation of SnB version 2.0. J. Appl. Cryst.* **32**, 120–124.

Weinzierl, J. E., Eisenberg, D. & Dickerson, R. E. (1969). *Refinement of protein phases with the Karle–Hauptman tangent fomula. Acta Cryst.* B**25**, 380–387.

White, P. & Woolfson, M. M. (1975). *The application of phase relationships to complex structures. VII. Magic integers. Acta Cryst.* A**31**, 53–56.

Wilkins, S. W., Varghese, J. N. & Lehmann, M. S. (1983). *Statistical geometry. I. A self-consistent approach to the crystallographic inversion problem based on information theory. Acta Cryst.* A**39**, 47–60.

Wilson, A. J. C. (1942). *Determination of absolute from relative X-ray intensity data. Nature (London)*, **150**, 151–152.

Wilson, K. S. (1978). *The application of MULTAN to the analysis of isomorphous derivatives in protein crystallography. Acta Cryst.* B**34**, 1599–1608.

Wolff, P. M. de & Bouman, J. (1954). *A fundamental set of structure factor inequalities. Acta Cryst.* **7**, 328–333.

Woolfson, M. M. (1958). *Crystal and molecular structure of p,p′-dimethoxybenzophenone by the direct probability method. Acta Cryst.* **11**, 277–283.

Woolfson, M. M. (1977). *On the application of phase relationships to complex structures. X. MAGLIN – a successor to MULTAN. Acta Cryst.* A**33**, 219–225.

Yao, J.-X. (1981). *On the application of phase relationships to complex structures. XVIII. RANTAN – random MULTAN. Acta Cryst.* A**37**, 642–664.

International Tables for Crystallography (2010). Vol. B, Chapter 2.3, pp. 244–281.

2.3. Patterson and molecular replacement techniques, and the use of noncrystallographic symmetry in phasing

By L. Tong, M. G. Rossmann and E. Arnold

2.3.1. Introduction

2.3.1.1. Background

Historically, the Patterson has been used in a variety of ways to effect the solutions of crystal structures. While some simple structures (Ketelaar & de Vries, 1939; Hughes, 1940; Speakman, 1949; Shoemaker *et al.*, 1950) were solved by direct analysis of Patterson syntheses, alternative methods have largely superseded this procedure. An early innovation was the heavy-atom method which depends on the location of a small number of relatively strong scatterers (Harker, 1936). Image-seeking methods and Patterson superposition techniques were first contemplated in the late 1930s (Wrinch, 1939) and applied sometime later (Beevers & Robertson, 1950; Clastre & Gay, 1950; Garrido, 1950*a*; Buerger, 1959). This experience provided the encouragement for computerized vector-search methods to locate individual atoms automatically (Mighell & Jacobson, 1963; Kraut, 1961; Hamilton, 1965; Simpson *et al.*, 1965) or to position known molecular fragments in unknown crystal structures (Nordman & Nakatsu, 1963; Huber, 1965). The Patterson function has been used extensively in conjunction with the isomorphous replacement method (Rossmann, 1960; Blow, 1958) or anomalous dispersion (Rossmann, 1961*a*) to determine the position of heavy-atom substitution. Pattersons have been used to detect the presence and relative orientation of multiple copies of a given chemical motif in the crystallographic asymmetric unit in the same or different crystals (Rossmann & Blow, 1962). Finally, the orientation and placement of known molecular structures ('molecular replacement') into unknown crystal structures can be accomplished *via* Patterson techniques.

The function, introduced by Patterson in 1934 (Patterson, 1934*a*,*b*), is a convolution of electron density with itself and may be defined as

$$P(\mathbf{u}) = \int_V \rho(\mathbf{x}) \cdot \rho(\mathbf{u}+\mathbf{x})\, \mathrm{d}\mathbf{x}, \qquad (2.3.1.1)$$

where $P(\mathbf{u})$ is the 'Patterson' function at $\mathbf{u}$, $\rho(\mathbf{x})$ is the crystal's periodic electron density and V is the volume of the unit cell. The Patterson function, or F^2 series, can be calculated directly from the experimentally derived X-ray intensities as

$$P(\mathbf{u}) = \frac{2}{V^2} \sum_{\mathbf{h}}^{\text{hemisphere}} |\mathbf{F}_\mathbf{h}|^2 \cos 2\pi\mathbf{h}\cdot\mathbf{u}. \qquad (2.3.1.2)$$

The derivation of (2.3.1.2) from (2.3.1.1) can be found in this volume (see Section 1.3.4.2.1.6) along with a discussion of the physical significance and symmetry of the Patterson function, although the principal properties will be restated here.

The Patterson can be considered to be a vector map of all the pairwise interactions between the atoms in a unit cell. The vectors in a Patterson correspond to vectors in the real (direct) crystal cell but translated to the Patterson origin. Their weights are proportional to the product of densities at the tips of the vectors in the real cell. The Patterson unit cell has the same size as the real crystal cell. The symmetry of the Patterson comprises the Laue point group of the crystal cell plus any additional lattice symmetry due to Bravais centring. The reduction of the real space group to the Laue symmetry is produced by the translation of all vectors to the Patterson origin and the introduction of a centre of symmetry. The latter is a consequence of the relationship between the vectors **AB** and **BA**. The Patterson symmetries for all 230 space groups are tabulated in *IT* A (2005).

An analysis of Patterson peaks can be obtained by considering N atoms with form factors f_i in the unit cell. Then

$$\mathbf{F}_\mathbf{h} = \sum_{i=1}^{N} f_i \exp(2\pi i \mathbf{h}\cdot\mathbf{x}_i).$$

Using Friedel's law,

$$\begin{aligned} |\mathbf{F}_\mathbf{h}|^2 &= \mathbf{F}_\mathbf{h}\cdot\mathbf{F}_\mathbf{h}^* \\ &= \left[\sum_{i=1}^{N} f_i \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_i)\right]\left[\sum_{j=1}^{N} f_j \exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)\right], \end{aligned}$$

which can be decomposed to

$$|\mathbf{F}_\mathbf{h}|^2 = \sum_{i=1}^{N} f_i^2 + \sum_{i\neq j}^{N}\sum_{j}^{N} f_i f_j \exp[2\pi i\mathbf{h}\cdot(\mathbf{x}_i-\mathbf{x}_j)]. \qquad (2.3.1.3)$$

On substituting (2.3.1.3) in (2.3.1.2), we see that the Patterson consists of the sum of N^2 total interactions of which N are of weight f_i^2 at the origin and $N(N-1)$ are of weight $f_i f_j$ at $\mathbf{x}_i - \mathbf{x}_j$.

The weight of a peak in a real cell is given by

$$w_i = \int_U \rho_i(\mathbf{x})\, \mathrm{d}\mathbf{x} = Z_i \text{ (the atomic number)},$$

where U is the volume of the atom i. By analogy, the weight of a peak in a Patterson (form factor $f_i f_j$) will be given by

$$w_{ij} = \int_U P_{ij}(\mathbf{u})\, \mathrm{d}\mathbf{u} = Z_i Z_j.$$

Although the maximum height of a peak will depend on the spread of the peak, it is reasonable to assume that heights of peaks in a Patterson are proportional to the products of the atomic numbers of the interacting atoms.

There are a total of N^2 interactions in a Patterson due to N atoms in the crystal cell. These can be represented as an $N\times N$ square matrix whose elements $\mathbf{u}_{ij}$, w_{ij} indicate the position and weight of the peak produced between atoms i and j (Table 2.3.1.1). The N vectors corresponding to the diagonal of this matrix are located at the Patterson origin and arise from the convolution of each atom with itself. This leaves $N(N-1)$ vectors whose locations depend on the relative positions of all of the atoms in the crystal cell and whose weights depend on the atom types related by the vector. Complete specification of the unique non-origin Patterson vectors requires description of only the $N(N-1)/2$ elements in either the upper or the lower triangle of this matrix, since the two sets of vectors represented by the two triangles are related by a centre of symmetry

Copyright © 2010 International Union of Crystallography

Table 2.3.1.1. *Matrix representation of Patterson peaks*

The $N \times N$ matrix represents the position $\mathbf{u}_{ij}$ and weights w_{ij} of atomic interactions in a Patterson arising from N atoms at $\mathbf{x}_i$ and weight w_i in the real cell.

	$\mathbf{x}_1, w_1$	$\mathbf{x}_2, w_2$	...	$\mathbf{x}_N, w_N$
$\mathbf{x}_1, w_1$	$\mathbf{u}_{11} = \mathbf{x}_1 - \mathbf{x}_1$, $w_{11} = w_1^2$	$\mathbf{u}_{12} = \mathbf{x}_1 - \mathbf{x}_2$, $w_{12} = w_1 w_2$	...	$\mathbf{u}_{1N} = \mathbf{x}_1 - \mathbf{x}_N$, $w_{1N} = w_1 w_N$
$\mathbf{x}_2, w_2$	$\mathbf{x}_2 - \mathbf{x}_1, w_2 w_1$	$0, w_2^2$	...	$\mathbf{x}_2 - \mathbf{x}_N, w_2 w_N$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{x}_N, w_N$	$\mathbf{x}_N - \mathbf{x}_1, w_N w_1$	$\mathbf{x}_N - \mathbf{x}_2, w_N w_2$	...	$0, w_N^2$

$[\mathbf{u}_{ij} \equiv \mathbf{x}_i - \mathbf{x}_j = -\mathbf{u}_{ij} \equiv -(\mathbf{x}_j - \mathbf{x}_i)]$. Patterson vector positions are usually represented as $\langle uvw \rangle$, where u, v and w are expressed as fractions of the Patterson cell axes.

2.3.1.2. Limits to the number of resolved vectors

If we assume a constant number of atoms per unit volume, the number of atoms N in a unit cell increases in direct proportion with the volume of the unit cell. Since the number of non-origin peaks in the Patterson function is $N(N-1)$ and the Patterson cell is the same size as the real cell, the problem of overlapping peaks in the Patterson function becomes severe as N increases. To make matters worse, the breadth of a Patterson peak is roughly equal to the sum of the breadth of the original atoms. The effective width of a Patterson peak will also increase with increasing thermal motion, although this effect can be artificially reduced by sharpening techniques. Naturally, a loss of attainable resolution at high scattering angles will affect the resolution of atomic peaks in the real cell as well as peaks in the Patterson cell. If U is the van der Waals volume per average atom, then the fraction of the cell occupied by atoms will be $f = NU/V$. Similarly, the fraction of the cell occupied by Patterson peaks will be $2UN(N-1)/V$ or $2f(N-1)$. With the reasonable assumption that $f \simeq 0.1$ for a typical organic crystal, then the cell can contain at most five atoms ($N \leq 5$) for there to be no overlap, other than by coincidence, of the peaks in the Patterson. As N increases there will occur a background of peaks on which are superimposed features related to systematic properties of the structure.

The contrast of selected Patterson peaks relative to the general background level can be enhanced by a variety of techniques. For instance, the presence of heavy atoms not only enhances the size of a relatively small number of peaks but ordinarily ensures a larger separation of the peaks due to the light-atom skeleton on which the heavy atoms are hung. That is, the factor f (above) is substantially reduced. Another example is the effect of systematic atomic arrangements (*e.g.* α-helices or aromatic rings) resulting in multiple peaks which stand out above the background. In the isomorphous replacement method, isomorphous difference Pattersons are utilized in which the contrast of the Patterson interactions between the heavy atoms is enhanced by removal of the predominant interactions which involve the rest of the structure.

2.3.1.3. Modifications: origin removal, sharpening etc.

A. L. Patterson, in his first in-depth exposition of his newly discovered F^2 series (Patterson, 1935), introduced the major modifications to the Patterson which are still in use today. He illustrated, with one-dimensional Fourier series, the techniques of removing the Patterson origin peak, sharpening the overall function and also removing peaks due to atoms in special positions. Each one of these modifications can improve the interpretability of Pattersons, especially those of simple structures. Whereas the recommended extent of such modifications is controversial (Buerger, 1966), most studies which utilize Patterson functions do incorporate some of these techniques [see, for example, Jacobson *et al.* (1961), Braun *et al.* (1969) and Nordman (1980*a*)]. Since Patterson's original work, other workers have suggested that the Patterson function itself might be modified; Fourier inversion of the modified Patterson then provides a new and perhaps more tractable set of structure factors (McLachlan & Harker, 1951; Simonov, 1965; Raman, 1966; Corfield & Rosenstein, 1966). Karle & Hauptman (1964) suggested that an improved set of structure factors could be obtained from an origin-removed Patterson modified such that it was everywhere non-negative and that Patterson density values less than a bonding distance from the origin were set to zero. Nixon (1978) was successful in solving a structure which had previously resisted solution by using a set of structure factors which had been obtained from a Patterson in which the largest peaks had been attenuated.

The N origin peaks [see expression (2.3.1.3)] may be removed from the Patterson by using coefficients

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{mod}}|^2 = |\mathbf{F}_{\mathbf{h}}|^2 - \sum_{i=1}^{N} f_i^2.$$

A Patterson function using these modified coefficients will retain all interatomic vectors. However, the observed structure factors $\mathbf{F}_{\mathbf{h}}$ must first be placed on an absolute scale (Wilson, 1942) in order to match the scattering-factor term. In practice, Patterson origins can also be removed by using coefficients

$$\left|\mathbf{F}_{\mathbf{h},\mathrm{mod}}\right|^2 = \left|\mathbf{F}_{\mathbf{h}}\right|^2 - \langle \left|\mathbf{F}_{\mathbf{h}}\right|^2 \rangle,$$

where $\langle \left|\mathbf{F}_{\mathbf{h}}\right|^2 \rangle$ is the average reflection intensity, usually calculated in several resolution shells. This formula has the advantage that the observed structure factors do not need to be on absolute scale.

Analogous to origin removal, the vector interactions due to atoms in known positions can also be removed from the Patterson function. Patterson showed that non-origin Patterson peaks arising from known atoms 1 and 2 may be removed by using the expression

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{mod}}|^2 = |\mathbf{F}_{\mathbf{h}}|^2 - \sum_{i=1}^{N} f_i^2 t_i^2 - 2 f_1 f_2 t_1 t_2 \cos 2\pi \mathbf{h} \cdot (\mathbf{x}_1 - \mathbf{x}_2),$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the positions of atoms 1 and 2 and t_1 and t_2 are their respective thermal correction factors. Using one-dimensional Fourier series, Patterson illustrated how interactions due to known atoms can obscure other information.

Patterson also introduced a means by which the peaks in a Patterson function may be artificially sharpened. He considered the effect of thermal motion on the broadening of electron-density peaks and consequently their Patterson peaks. He suggested that the F^2 coefficients could be corrected for thermal effects by simulating the atoms as point scatterers and proposed using a modified set of coefficients

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}}|^2 = |\mathbf{F}_{\mathbf{h}}|^2 / \bar{f}^2,$$

where $\bar{f}$, the average scattering factor per electron, is given by

$$\bar{f} = \sum_{i=1}^{N} f_i \Big/ \sum_{i=1}^{N} Z_i.$$

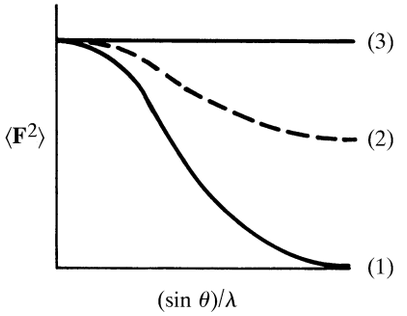


Fig. 2.3.1.1. Effect of 'sharpening' Patterson coefficients. (1) shows a mean distribution of $|\mathbf{F}|^2$ values with resolution, $(\sin\theta)/\lambda$. The normal decline of this curve is due to increasing destructive interference between different portions within diffuse atoms at larger Bragg angles. (2) shows the distribution of 'sharpened' coefficients. (3) shows the theoretical distribution of $|\mathbf{F}|^2$ produced by a point-atom structure. To represent such a structure with a Fourier series would require an infinite series in order to avoid large errors caused by truncation.

A common formulation for this type of sharpening expresses the atomic scattering factors at a given angle in terms of an overall isotropic thermal parameter B as

$$f(s) = f_0 \exp(-Bs^2).$$

The Patterson coefficients then become

$$\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}} = \frac{Z_{\mathrm{total}}}{\sum_{i=l}^{N} f(s)} \mathbf{F}_{\mathbf{h}}.$$

The normalized structure factors, E (see Chapter 2.2), which are used in crystallographic direct methods, are also a common source of sharpened Patterson coefficients $(E^2 - 1)$. Although the centre positions and total contents of Patterson peaks are unaltered by sharpening, the resolution of individual peaks may be enhanced. The degree of sharpening can be controlled by adjusting the size of the assumed B factor; Lipson & Cochran (1966, pp. 165–170) analysed the effect of such a choice on Patterson peak shape.

All methods of sharpening Patterson coefficients aim at producing a point atomic representation of the unit cell. In this quest, the high-resolution terms are enhanced (Fig. 2.3.1.1). Unfortunately, this procedure must also produce a serious Fourier truncation error which will be seen as large ripples about each Patterson peak (Gibbs, 1898). Consequently, various techniques have been used (mostly unsuccessfully) in an attempt to balance the advantages of sharpening with the disadvantages of truncation errors.

Schomaker and Shoemaker [unpublished; see Lipson & Cochran (1966, p. 168)] used a function

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}}|^2 = \frac{|\mathbf{F}_{\mathbf{h}}|^2}{\bar{f}^2} s^2 \exp\left[-\frac{\pi^2}{p} s^2\right],$$

in which s is the length of the scattering vector, to produce a Patterson synthesis which is less sensitive to errors in the low-order terms. Jacobson *et al.* (1961) used a similar function,

$$|\mathbf{F}_{\mathbf{h},\,\mathrm{sharp}}|^2 = \frac{|\mathbf{F}_{\mathbf{h}}|^2}{\bar{f}^2} (k + s^2) \exp\left[-\frac{\pi}{p} s^2\right],$$

which they rationalize as the sum of a scaled exponentially sharpened Patterson and a gradient Patterson function (the value of k was empirically chosen as $\frac{2}{3}$). This approach was subsequently further developed and generalized by Wunderlich (1965).

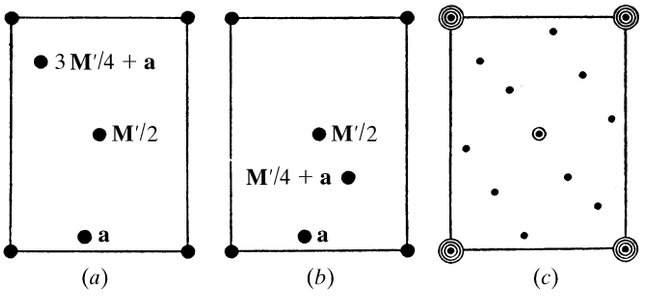


Fig. 2.3.1.2. (*c*) The point Patterson of the two homometric structures in (*a*) and (*b*). The latter are constructed by taking points at $\mathbf{a}$ and $\frac{1}{2}\mathbf{M}'$, where $\mathbf{M}'$ is a cell diagonal, and adding a third point which is (*a*) at $\frac{3}{4}\mathbf{M}' + \mathbf{a}$ or (*b*) at $\frac{1}{4}\mathbf{M}' + \mathbf{a}$. [Reprinted with permission from Patterson (1944).]

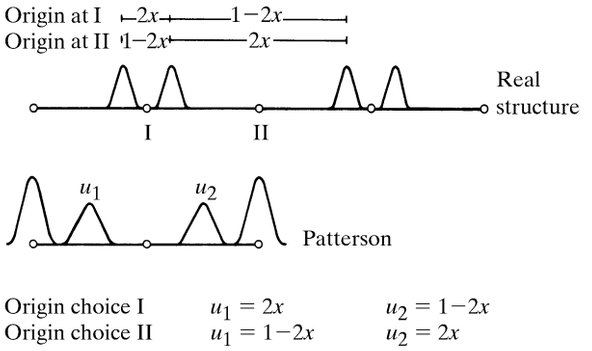
2.3.1.4. Homometric structures and the uniqueness of structure solutions; enantiomorphic solutions

Interpretation of any Patterson requires some assumption, such as the existence of discrete atoms. A complete interpretation might also require an assumption of the number of atoms and may require other external information (*e.g.* bond lengths, bond angles, van der Waals separations, hydrogen bonding, positive density *etc.*). To what extent is the solution of a Patterson function unique? Clearly the greater is the supply of external information, the fewer will be the number of possible solutions. Other constraints on the significance of a Patterson include the error involved in measuring the observed coefficients and the resolution limit to which they have been observed. The larger the error, the larger the number of solutions. When the error on the amplitudes is infinite, it is only the other physical constraints, such as packing, which limit the structural solutions. Alternative solutions of the same Patterson are known as homometric structures.

During their investigation of the mineral bixbyite, Pauling & Shappell (1930) made the disturbing observation that there were two solutions to the structure, with different arrangements of atoms, which yielded the same set of calculated intensities. Specifically, atoms occupying equipoint set 24*d* in space group $I(2_1/a)\bar{3}$ can be placed at either $x, 0, \frac{1}{4}$ or $-x, 0, \frac{1}{4}$ without changing the calculated intensities. Yet the two structures were not chemically equivalent. These authors resolved the ambiguity by placing the oxygen atoms in question at positions which gave the most acceptable bonding distances with the rest of the structure.

Patterson interpreted the above ambiguity in terms of the F^2 series: the distance vector sets or Patterson functions of the two structures were the same since each yielded the same calculated intensities (Patterson, 1939). He defined such a pair of structures a homometric pair and called the degenerate vector set which they produced a homometric set. Patterson went on to investigate the likelihood of occurrence of homometric structures and, indeed, devoted a great deal of his time to this matter. He also developed algebraic formalisms for examining the occurrence of homometric pairs and multiplets in selected one-dimensional sets of points, such as cyclotomic sets, and also sets of points along a line (Patterson, 1944). Some simple homometric pairs are illustrated in Fig. 2.3.1.2.

Drawing heavily from Patterson's inquiries into the structural uniqueness allowed by the diffraction data, Hosemann, Bagchi and others have given formal definitions of the different types of homometric structures (Hosemann & Bagchi, 1954). They suggested a classification divided into pseudohomometric structures and homomorphs, and used an integral equation representing a convolution operation to express their examples of finite homometric structures. Other workers have chosen various



Origin choice I	$u_1 = 2x$	$u_2 = 1-2x$
Origin choice II	$u_1 = 1-2x$	$u_2 = 2x$

Fig. 2.3.2.1. Origin selection in the interpretation of a Patterson of a one-dimensional centrosymmetric structure.

means for describing homometric structures [Buerger (1959, pp. 41–50), Menzer (1949), Bullough (1961, 1964), Hoppe (1962)].

Since a Patterson function is centrosymmetric, the Pattersons of a crystal structure and of its mirror image are identical. Thus the enantiomeric ambiguity present in noncentrosymmetric crystal structures cannot be overcome by using the Patterson alone and represents a special case of homometric structures. Assignment of the correct enantiomorph in a crystal structure analysis is generally not possible unless a recognizable fragment of known chirality emerges (*e.g.* L-amino acids in proteins, D-riboses in nucleic acids, the known framework of steroids and other natural products, the right-handed twist of α-helices, the left-handed twist of successive strands in a β-sheet, the fold of a known protein subunit *etc.*) or anomalous-scattering information is available and can be used to resolve the ambiguity (Bijvoet *et al.*, 1951).

It is frequently necessary to select arbitrarily one enantiomorph over another in the early stages of a structure solution. Structure-factor phases calculated from a single heavy atom in space group $P1$, $P2$ or $P2_1$ (for instance) will be centrosymmetric and both enantiomorphs will be present in Fourier calculations based on these phases. In other space groups (*e.g.* $P2_12_12_1$), the selected heavy atom is likely to be near one of the planes containing the 2_1 axes and thus produce a weaker 'ghost' image of the opposite enantiomorph. The mixture of the two overlapped enantiomorphic solutions can cause interpretive difficulties. As the structure solution progresses, the 'ghosts' are exorcized owing to the dominance of the chosen enantiomorph in the phases.

2.3.1.5. The Patterson synthesis of the second kind

Patterson also defined a second, less well known, function (Patterson, 1949) as

$$P_{\pm}(\mathbf{u}) = \int \rho(\mathbf{u}+\mathbf{x}) \cdot \rho(\mathbf{u}-\mathbf{x})\ \mathrm{d}\mathbf{x}$$
$$= \frac{2}{V^2} \sum_{\mathbf{h}}^{\text{hemisphere}} F_{\mathbf{h}}^2 \cos(2\pi 2\mathbf{h}\cdot\mathbf{u} - 2\alpha_{\mathbf{h}}).$$

This function can be computed directly only for centrosymmetric structures. It can be calculated for noncentrosymmetric structures when the phase angles are known or assumed. It will represent the degree to which the known or assumed structure has a centre of symmetry at $\mathbf{u}$. That is, the product of the density at $\mathbf{u}+\mathbf{x}$ and $\mathbf{u}-\mathbf{x}$ is large when integrated over all values $\mathbf{x}$ within the unit cell. Since atoms themselves have a centre of symmetry, the function will contain peaks at each atomic site roughly proportional in height to the square of the number of electrons in each atom plus peaks at the midpoint between atoms proportional in height to the product of the electrons in each atom. Although this function has not been found very useful in practice, it is useful for demonstrating the presence of weak enantiomorphic images in a given tentative structure determination.

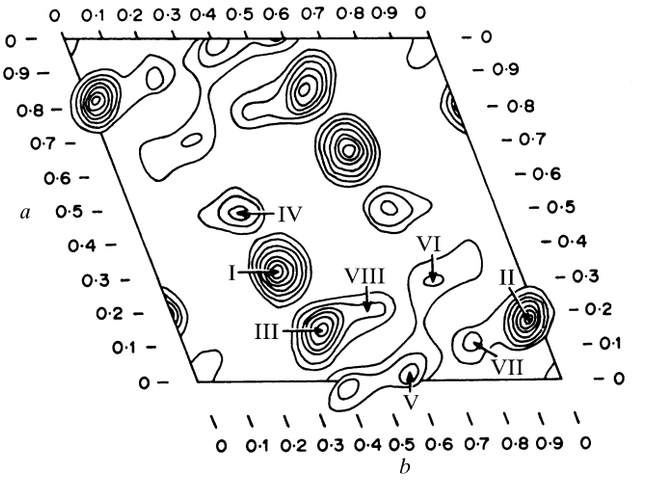


Fig. 2.3.2.2. The *c*-axis projection of cuprous chloride azomethane complex ($C_2H_6Cl_2Cu_2N_2$). The space group is $P\bar{1}$ with one molecule per unit cell. [Adapted from and reprinted with permission from Woolfson (1970, p. 321).]

2.3.2. Interpretation of Patterson maps

2.3.2.1. Simple solutions in the triclinic cell. Selection of the origin

A hypothetical one-dimensional centrosymmetric crystal structure containing an atom at x and at $-x$ and the corresponding Patterson is illustrated in Fig. 2.3.2.1. There are two different centres of symmetry which may be chosen as convenient origins. If the atoms are of equal weight, we expect Patterson vectors at positions $u = \pm 2x$ with weights equal to half the origin peak. There are two symmetry-related peaks, u_1 and u_2 (Fig. 2.3.2.1) in the Patterson. It is an arbitrary choice whether $u_1 = 2x$ or $u_2 = 2x$. This choice is equivalent to selecting the origin at the centre of symmetry I or II in the real structure (Fig. 2.3.2.1). Similarly in a three-dimensional $P\bar{1}$ cell, the Patterson will contain peaks at $\langle uvw\rangle$ which can be used to solve for the atom coordinates $\pm\langle 2x, 2y, 2z\rangle$. Solving for the same coordinates by starting from symmetric representations of the same vector will lead to alternate origin choices. For example, use of $\langle 1+u, 1+v, w\rangle$ will lead to translating the origin by $(+\frac{1}{2}, +\frac{1}{2}, 0)$ relative to the solution based on $\langle uvw\rangle$. There are eight distinct inversion centres in $P\bar{1}$, each one of which represents a valid origin choice. Although any choice of origin would be allowable, an inversion centre is convenient because then the structure factors are all real. Typically, one of the vector peaks closest to the Patterson origin is selected to start the solution, usually in the calculated asymmetric unit of the Patterson. Care must be exercised in selecting the same origin for all atomic positions by considering cross-vectors between atoms.

Examine, for example, the *c*-axis Patterson projection of a cuprous chloride azomethane complex ($C_2H_6Cl_2Cu_2N_2$) in $P\bar{1}$ as shown in Fig. 2.3.2.2. The largest Patterson peaks should correspond to vectors arising from Cu ($Z = 29$) and Cl ($Z = 17$) atoms. There will be copper atoms at $\mathbf{x}_{Cu}(x_{Cu}, y_{Cu})$ and $-\mathbf{x}_{Cu}(-x_{Cu}, -y_{Cu})$ as well as chlorine atoms at analogous positions. The interaction matrix is

	$\mathbf{x}_{Cu}$, 29	$\mathbf{x}_{Cl}$, 17	$-\mathbf{x}_{Cu}$, 29	$-\mathbf{x}_{Cl}$, 17
$\mathbf{x}_{Cu}$, 29	0, 841	$\mathbf{x}_{Cu}-\mathbf{x}_{Cl}$, 493	$2\mathbf{x}_{Cu}$, 841	$\mathbf{x}_{Cu}+\mathbf{x}_{Cl}$, 493
$\mathbf{x}_{Cl}$, 17		0, 289	$\mathbf{x}_{Cl}+\mathbf{x}_{Cu}$, 493	$2\mathbf{x}_{Cl}$, 289
$-\mathbf{x}_{Cu}$, 29			0, 841	$\mathbf{x}_{Cu}-\mathbf{x}_{Cl}$, 493
$-\mathbf{x}_{Cl}$, 17				0, 289

Table 2.3.2.1. *Coordinates of Patterson peaks for $C_2H_6Cl_2Cu_2N_2$ projection*

Height	u	v	Number in diagram (Fig. 2.3.2.2)
7	0.33	0.34	I
7	0.18	0.97	II
6	0.16	0.40	III
3	0.49	0.29	IV
3	0.02	0.59	V
2	0.30	0.75	VI
2	0.12	0.79	VII

which shows that the Patterson should contain the following types of vectors:

Position	Weight	Multiplicity	Total weight
$2\mathbf{x}_{\rm Cu}$	841	1	841
$2\mathbf{x}_{\rm Cl}$	289	1	289
$\mathbf{x}_{\rm Cu}-\mathbf{x}_{\rm Cl}$	493	2	986
$\mathbf{x}_{\rm Cu}+\mathbf{x}_{\rm Cl}$	493	2	986

The coordinates of the largest Patterson peaks are given in Table 2.3.2.1 for an asymmetric half of the cell chosen to span $0 \rightarrow \frac{1}{2}$ in u and $0 \rightarrow 1$ in v. Since the three largest peaks are in the same ratio (7:7:6) as the three largest expected vector types (986:986:841), it is reasonable to assume that peak III corresponds to the copper–copper interaction at $2\mathbf{x}_{\rm Cu}$. Hence, $x_{\rm Cu} = 0.08$ and $y_{\rm Cu} = 0.20$. Peaks I and II should be due to the double-weight Cu–Cl vectors at $\mathbf{x}_{\rm Cu} - \mathbf{x}_{\rm Cl}$ and $\mathbf{x}_{\rm Cu} + \mathbf{x}_{\rm Cl}$. Now suppose that peak I is at position $\mathbf{x}_{\rm Cu} + \mathbf{x}_{\rm Cl}$, then $x_{\rm Cl} = 0.25$ and $y_{\rm Cl} = 0.14$. Peak II should now check out as the remaining double-weight Cu–Cl interaction at $\mathbf{x}_{\rm Cu} - \mathbf{x}_{\rm Cl}$. Indeed, $\mathbf{x}_{\rm Cu} - \mathbf{x}_{\rm Cl} = \langle -0.17, 0.06\rangle = -\langle 0.17, -0.06\rangle$ which agrees tolerably well with the position of peak II. The chlorine position also predicts the position of a peak at $2\mathbf{x}_{\rm Cl}$ with weight 289; peak IV confirms the chlorine assignment. In fact, this Patterson can be solved also for the lighter nitrogen- and carbon-atom positions which account for the remainder of the vectors listed in Table 2.3.2.1. However, the simplest way to complete the structure determination is probably to compute a Fourier synthesis using phases calculated from the heavier copper and chlorine positions.

Consider now a real cell with M crystallographic asymmetric units, each of which contains N atoms. Let us define $\mathbf{x}_{mn}$, the position of the nth atom in the mth crystallographic unit, by

$$\mathbf{x}_{mn} = [\boldsymbol{T}_m]\mathbf{x}_{1n} + \mathbf{t}_m,$$

where $[\boldsymbol{T}_m]$ and $\mathbf{t}_m$ are the rotation matrix and translation vector, respectively, for the mth crystallographic symmetry operator. The Patterson of this crystal will contain vector peaks which arise from atoms interacting with other atoms both in the same and in different crystallographic asymmetric units. The set of $(MN)^2$ Patterson vector interactions for this crystal is represented in a matrix in Table 2.3.2.2. Upon dissection of this diagram we see that there are MN origin vectors, $M[(N-1)N]$ vectors from atom interactions with other atoms in the same crystallographic asymmetric unit and $[M(M-1)]N^2$ vectors involving atoms in separate asymmetric units. Often a number of vectors of special significance relating symmetry-equivalent atoms emerge from this milieu of Patterson vectors and such 'Harker vectors' constitute the subject of the next section.

2.3.2.2. Harker sections

Soon after Patterson introduced the F^2 series, Harker (1936) recognized that many types of crystallographic symmetry result in a concentration of vectors at characteristic locations in the Patterson. Specifically, he showed that atoms related by rotation axes produce vectors in characteristic planes of the Patterson, and that atoms related by mirror planes or reflection glide planes produce vectors on characteristic lines. Similarly, noncrystallographic symmetry operators produce analogous concentrations of vectors. Harker showed how special sections through a three-dimensional function could be computed using one- or two-dimensional summations. With the advent of powerful computers, it is usual to calculate a full three-dimensional Patterson synthesis. Nevertheless, 'Harker' planes or lines are often the starting point for a structure determination. It should, however, be noted that non-Harker vectors (those not due to interactions between symmetry-related atoms) can appear by coincidence in a Harker section. Table 2.3.2.3 shows the position in a Patterson of Harker planes and lines produced by all types of crystallographic symmetry operators.

Buerger (1946) noted that Harker sections can be helpful in space-group determination. Concentrations of vectors in appropriate regions of the Patterson should be diagnostic for the presence of some symmetry elements. This is particularly useful where these elements (such as mirror planes) are not directly detected by systematic absences.

Buerger also developed a systematic method of interpreting Harker peaks which he called implication theory [Buerger (1959, Chapter 7)].

Table 2.3.2.2. *Square matrix representation of vector interactions in a Patterson of a crystal with M crystallographic asymmetric units each containing N atoms*

Peak positions $\mathbf{u}_{m1n1,\,m2n2}$ correspond to vectors between the atoms $\mathbf{x}_{m1n1}$ and $\mathbf{x}_{m2n2}$ where $\mathbf{x}_{mn}$ is the nth atom in the mth crystallographic asymmetric unit. The corresponding weights are $w_{n1}w_{n2}$. The outlined blocks I1 and IM represent vector interactions between atoms in the same crystallographic asymmetric units (there are M such blocks). The off-diagonal blocks IIM1 and II1M represent vector interactions between atoms in crystal asymmetric units 1 and M; there are $M(M-1)$ blocks of this type. The significance of diagonal elements of block IIM1 is that they represent Harker-type interactions between symmetry-equivalent atoms (see Section 2.3.2.2).

	$\mathbf{x}_{11}, w_1$	$\mathbf{x}_{12}, w_2$	...	$\mathbf{x}_{1N}, w_N$	...	$\mathbf{x}_{M1}, w_1$	$\mathbf{x}_{M2}, w_2$	...	$\mathbf{x}_{MN}, w_N$
$\mathbf{x}_{11}, w_1$	$0, w_1^2$	$\mathbf{u}_{11,12}, w_1w_2$	...	$\mathbf{u}_{11,1N}, w_1w_N$	...				
$\mathbf{x}_{12}, w_2$		$0, w_2^2$	...	$\mathbf{u}_{12,1N}, w_2w_N$					
⋮		⋱		⋮	...				
$\mathbf{x}_{1N}, w_N$				$0, w_N^2$					
	Block I1					Block II1M			
⋮	⋮				⋱		⋮		
$\mathbf{x}_{M1}, w_1$	$\mathbf{u}_{M1,11}, w_1^2$	$\mathbf{u}_{M1,12}, w_1w_2$							
$\mathbf{x}_{M2}, w_2$	$\mathbf{u}_{M2,11}, w_2w_1$	$\mathbf{u}_{M2,12}, w_2^2$							
⋮		⋱			...				
$\mathbf{x}_{MN}, w_N$				$\mathbf{u}_{MN,1N}, w_N^2$					
	Block IIM1					Block IM			

Table 2.3.2.3. *Position of Harker sections within a Patterson*

Symmetry element	Form of $P(x, y, z)$
(*a*) Harker planes	
Axes parallel to the *b* axis:	
(i) 2	$P(x, 0, z)$
(ii) 2_1	$P(x, \frac{1}{2}, z)$
Axes parallel to the *c* axis:	
(i) 2, 3, $\bar{3}$, 4, $\bar{4}$, 6, $\bar{6}$	$P(x, y, 0)$
(ii) 2_1, 4_2, 6_3	$P(x, y, \frac{1}{2})$
(iii) 3_1, 3_2, 6_2, 6_4	$P(x, y, \frac{1}{3})$
(iv) 4_1, 4_3	$P(x, y, \frac{1}{4})$
(v) 6_1, 6_5	$P(x, y, \frac{1}{6})$
(*b*) Harker lines	
Planes perpendicular to the *b* axis:	
(i) Reflection planes	$P(0, y, 0)$
(ii) Glide plane, glide $= \frac{1}{2}a$	$P(\frac{1}{2}, y, 0)$
(iii) Glide plane, glide $= \frac{1}{2}c$	$P(0, y, \frac{1}{2})$
(iv) Glide plane, glide $= \frac{1}{2}(a + c)$	$P(\frac{1}{2}, y, \frac{1}{2})$
(v) Glide plane, glide $= \frac{1}{4}(a + c)$	$P(\frac{1}{4}, y, \frac{1}{4})$
(vi) Glide plane, glide $= \frac{1}{4}(3a + c)$	$P(\frac{3}{4}, y, \frac{1}{4})$
(*c*) Special Harker planes	
Axes parallel to or containing body diagonal (111), valid for cubic space groups only:	
	Equation of plane $lx + my + nz - p = 0$
(i) 3	$l = m = n = \cos 54.73561^\circ = 0.57735$, $p = 0$
(ii) 3_1	$l = m = n = \cos 54.73561^\circ = 0.57735$, $p = \sqrt{3}/3$
Rhombohedral threefold axes produce analogous Harker planes whose description will depend on the interaxial angle.	

2.3.2.3. *Finding heavy atoms*

The previous two sections have developed some of the useful mechanics for interpreting Pattersons. In this section, we will consider finding heavy-atom positions, in the presence of numerous light atoms, from Patterson maps. The feasibility of structure solution by the heavy-atom method depends on a number of factors which include the relative size of the heavy atom and the extent and quality of the data. A useful rule of thumb is that the ratio

$$r = \frac{\sum_{\text{heavy}} Z^2}{\sum_{\text{light}} Z^2}$$

should be near unity if the heavy atom is to provide useful starting phase information (Z is the atomic number of an atom). The condition that $r > 1$ normally guarantees interpretability of the Patterson function in terms of the heavy-atom positions. This 'rule', arising from the work of Luzzati (1953), Woolfson (1956), Sim (1961) and others, is not inviolable; many ambitious determinations have been accomplished *via* the heavy-atom method for which r was well below 1.0. An outstanding example is vitamin B_{12} with formula $C_{62}H_{88}CoO_{14}P$, which gave an $r = 0.14$ for the cobalt atom alone (Hodgkin *et al.*, 1957). One factor contributing to the success of such a determination is that the relative scattering power of Co is enhanced for higher scattering angles. Thus, the ratio, r, provides a conservative estimate. If the value of r is well above 1.0, the initial easier interpretation of the Patterson will come at the expense of poorly defined parameters of the lighter atoms.

A general strategy for determining heavy atoms from the Patterson usually involves the following steps.

(1) List the number and type of atoms in the cell.

(2) Construct the interaction matrix for the heaviest atoms to predict the positions and weights of the largest Patterson vectors. Group recurrent vectors and notice vectors with special properties, such as Harker vectors.

(3) Compute the Patterson using any desired modifications. Placing the map on an absolute scale [$P(000) = \sum Z^2$] is convenient but not necessary.

(4) Examine Harker sections and derive trial atom coordinates from vector positions.

(5) Check the trial coordinates using other vectors in the predicted set. Correlate enantiomorphic choice and origin choice for independent sites.

(6) Include the next-heaviest atoms in the interpretation if possible. In particular, use the cross-vectors with the heaviest atoms.

(7) Use the best heavy-atom model to initiate phasing.

Detailed and instructive examples of using Pattersons to find heavy-atom positions are found in almost every textbook on crystal structure analysis [see, for example, Buerger (1959), Lipson & Cochran (1966) and Stout & Jensen (1968)].

The determination of the crystal structure of cholesteryl iodide by Carlisle & Crowfoot (1945) provides an example of using the Patterson function to locate heavy atoms. There were two molecules, each of formula $C_{27}H_{45}I$, in the $P2_1$ unit cell. The ratio $r = 2.8$ is clearly well over the optimal value of unity. The $P(x, z)$ Patterson projection showed one dominant peak at $\langle 0.434, 0.084 \rangle$ in the asymmetric unit. The equivalent positions for $P2_1$ require that an iodine atom at x_I, y_I, z_I generates another at $-x_I, \frac{1}{2} + y_I, z_I$ and thus produces a Patterson peak at $\langle 2x_I, \frac{1}{2}, 2z_I \rangle$. The iodine position was therefore determined as 0.217, 0.042. The y coordinate of the iodine is arbitrary for $P2_1$ yet the value of $y_I = 0.25$ is convenient, since an inversion centre in the two-atom iodine structure is then exactly at the origin, making all calculated phases 0 or π. Although the presence of this extra symmetry caused some initial difficulties in the interpretation of the steroid backbone, Carlisle and Crowfoot successfully separated the enantiomorphic images. Owing to the presence of the perhaps *too heavy* iodine atom, however, the structure of the carbon skeleton could not be defined very precisely. Nevertheless, all critical stereochemical details were adequately illuminated by this determination. In the cholesteryl iodide example, a number of different yet equivalent origins could have been selected. Alternative origin choices include all combinations of $x \pm \frac{1}{2}$ and $z \pm \frac{1}{2}$.

A further example of using the Patterson to find heavy atoms will be provided in Section 2.3.5.2 on solving for heavy atoms in the presence of noncrystallographic symmetry.

2.3.2.4. *Superposition methods. Image detection*

As early as 1939, Wrinch (1939) showed that it was possible, in principle, to recover a fundamental set of points from the vector map of that set. Unlike the Harker–Buerger implication theory (Buerger, 1946), the method that Wrinch suggested was capable of using all the vectors in a three-dimensional set, not those restricted to special lines or sections. Wrinch's ideas were developed for vector sets of points, however, and could not be directly applied to real, heavily overlapped Pattersons of a complex structure. These ideas seem to have lain dormant until the early 1950s when a number of independent investigators developed superposition methods (Beevers & Robertson, 1950; Clastre & Gay, 1950; Garrido, 1950*a*; Buerger, 1950*a*).

A Patterson can be considered as a sum of images of a molecule as seen, in turn, for each atom placed on the origin (Fig. 2.3.2.3). Thus, the deconvolution of a Patterson could proceed by superimposing each image of the molecule obtained onto the others by translating the Patterson origin to each imaging atom.

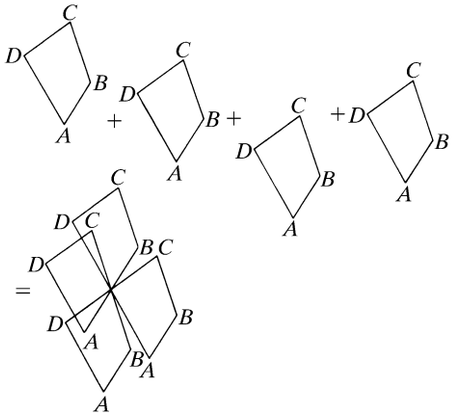


Fig. 2.3.2.3. Atoms *ABCD*, arranged as a quadrilateral, generate a Patterson which is the sum of the images of the quadrilateral when each atom is placed on the origin in turn.

For instance, let us take a molecule consisting of four atoms *ABCD* arranged in the form of a quadrilateral (Fig. 2.3.2.3). Then the Patterson consists of the images of four identical quadrilaterals with atoms *A*, *B*, *C* and *D* placed on the origin in turn. The Pattersons can then be deconvoluted by superimposing two of these Pattersons after translating these (without rotation) by, for instance, the vector **AB**. A further improvement could be obtained by superimposing a third Patterson translated by **AC**. This would have the additional advantage in that *ABC* is a noncentrosymmetric arrangement and, therefore, selects the enantiomorph corresponding to the hand of the atomic arrangement *ABC* [*cf.* Buerger (1951, 1959)].

A basic problem is that knowledge of the vectors **AB** and **AC** also implies some knowledge of the structure at a time when the structure is not yet known. In practice 'good-looking' peaks, estimated to be single peaks by assessing the absolute scale of the structure amplitudes with Wilson statistics, can be assumed to be the result of single interatomic vectors within a molecule. Superposition can then proceed and the result can be inspected for reasonable chemical sense. As many apparently single peaks can be tried systematically using a computer, this technique is useful for selecting and testing a series of reasonable Patterson interpretations (Jacobson *et al.*, 1961).

Three major methods have been used for the detection of molecular images of superimposed Pattersons. These are the sum, product and minimum 'image seeking' functions (Raman & Lipscomb, 1961). The concept of the sum function is to add the images where they superimpose, whereas elsewhere the summed Pattersons will have a lower value owing to lack of image superposition. Therefore, the sum function determines the average Patterson density for superimposed images, and is represented analytically as

$$S(\mathbf{x}) = \sum_{i=1}^{N} P(\mathbf{x} + \mathbf{u}_i),$$

where $S(\mathbf{x})$ is the sum function at $\mathbf{x}$ given by the superposition of the ith Patterson translated by $\mathbf{u}_i$, or

$$S(\mathbf{x}) = \sum_{\mathbf{h}} \left\{ F_{\mathbf{h}}^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) \left[\sum_{i=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{u}_i) \right] \right\}.$$

Setting

$$m \exp(i\varphi_{\mathbf{h}}) = \sum_{i=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{u}_i)$$

(m and $\varphi_{\mathbf{h}}$ can be calculated from the translational vectors used for the superposition),

$$S(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}}^2 m \exp(2\pi i \mathbf{h} \cdot \mathbf{x} + \varphi_{\mathbf{h}}).$$

Thus, the sum function is equivalent to a weighted 'heavy atom' method based on the known atoms assumed by the superposition translation vectors.

The product function is somewhat more vigorous in that the images are enhanced by the product. If an image is superimposed on no image, then the product should be zero.

The product function can be expressed as

$$Pr(\mathbf{x}) = \prod_{i=1}^{N} P(\mathbf{x} + \mathbf{u}_i).$$

When $N = 2$ ($\mathbf{h}$ and $\mathbf{p}$ are sets of Miller indices),

$$Pr(\mathbf{x}) = \sum_{\mathbf{h}} \sum_{\mathbf{p}} F_{\mathbf{h}}^2 F_{\mathbf{p}}^2 \exp[2\pi i(\mathbf{h} + \mathbf{p}) \cdot \mathbf{x}] \times \exp[2\pi i(\mathbf{h} \cdot \mathbf{u}_i + \mathbf{p} \cdot \mathbf{u}_i)].$$

Successive superpositions using the product functions will quickly be dominated by a few terms with very large coefficients.

Finally, the minimum function is a clever invention of Buerger (Buerger, 1950*b*, 1951, 1953*a*,*b*,*c*; Taylor, 1953; Rogers, 1951). If a superposition is correct then each Patterson must represent an image of the structure. Whenever there are two or more images that intersect in the Patterson, the Patterson density will be greater than a single image. When two different images are superimposed, it is a reasonable hope that at least one of these is a single image. Thus by taking the value of that Patterson which is the minimum, it should be possible to select a single image and eliminate noise from interfering images as far as possible. Although the minimum function is perhaps the most powerful algorithm for image selection of well sharpened Pattersons, it is not readily amenable to Fourier representation.

The minimum function was conceived on the basis of selecting positive images on a near-zero background. If it were desired to select negative images [*e.g.* the $(\mathbf{F}_1 - \mathbf{F}_2)^2$ correlation function discussed in Section 2.3.3.4], then it would be necessary to use a maximum function. In fact, normally, an image has finite volume with varying density. Thus, some modification of the minimum function is necessary in those cases where the image is large compared to the volume of the unit cell, as in low-resolution protein structures (Rossmann, 1961*b*). Nordman (1966) used the average of the Patterson values of the lowest 10 to 20 per cent of the vectors in comparing Pattersons with hypothetical point Pattersons. A similar criterion was used by High & Kraut (1966).

Image-seeking methods using Patterson superposition have been used extensively (Beevers & Robertson, 1950; Garrido, 1950*b*; Robertson, 1951). For a review the reader is referred to *Vector Space* (Buerger, 1959) and a paper by Fridrichsons & Mathieson (1962). However, with the advent of computerized direct methods (see Chapter 2.2), such techniques are no longer popular. Nevertheless, they provide the conceptual framework for the rotation and translation functions (see Sections 2.3.6 and 2.3.7).

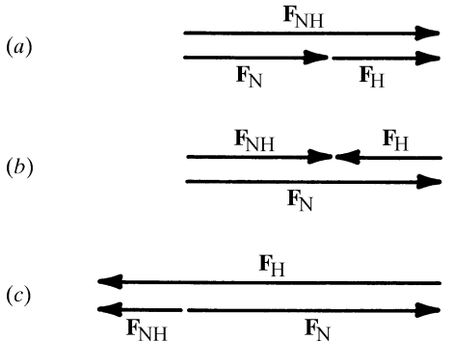


Fig. 2.3.3.1. Three different cases which can occur in the relation of the native, $\mathbf{F}_N$, and heavy-atom derivative, $\mathbf{F}_{NH}$, structure factors for centrosymmetric reflections. $\mathbf{F}_N$ is assumed to have a phase of 0, although analogous diagrams could be drawn when $\mathbf{F}_N$ has a phase of π. The crossover situation in (*c*) is clearly rare if the heavy-atom substitution is small compared to the native molecule, and can in general be neglected.

2.3.2.5. Systematic computerized Patterson vector-search procedures. Looking for rigid bodies

The power of the modern digital computer has enabled rapid access to the large number of Patterson density values which can serve as a lookup table for systematic vector-search procedures. In the late 1950s, investigators began to use systematic searches for the placement of single atoms, of known chemical groups or fragments and of entire known structures. Methods for locating single atoms were developed and called variously: vector verification (Mighell & Jacobson, 1963), symmetry minimum function (Kraut, 1961; Simpson *et al*., 1965; Corfield & Rosenstein, 1966) and consistency functions (Hamilton, 1965). Patterson superposition techniques using stored function values were often used to image the structure from the known portion. In such single-site search procedures, single atoms are placed at all possible positions in a crystal, using a search grid of the same fineness as for the stored Patterson function, preferably about one-third of the resolution of the Patterson map. Solutions are gauged to be acceptable if all expected vectors due to symmetry-related atoms are observed above a specified threshold in the Patterson.

Systematic computerized Patterson search procedures for orienting and positioning known molecular fragments were also developed in the early 1960s. These hierarchical procedures rely on first using the 'self'-vectors which depend only on the orientation of a molecular fragment. A search for the position of the fragment relative to the crystal symmetry elements then uses the cross-vectors between molecules (see Sections 2.3.6 and 2.3.7). Nordman constructed a weighted point representation of the predicted vector set for a fragment (Nordman & Nakatsu, 1963; Nordman, 1966) and successfully solved the structure of a number of complex alkaloids. Huber (1965) used the convolution molecule method of Hoppe (1957*a*) in three dimensions to solve a number of natural-product structures, including steroids. Various program systems have used Patterson search methods operating in real space to solve complex structures (Braun *et al*., 1969; Egert, 1983).

Others have used reciprocal-space procedures for locating known fragments. Tollin & Cochran (1964) developed a procedure for determining the orientation of planar groups by searching for origin-containing planes of high density in the Patterson. General procedures using reciprocal-space representations for determining rotation and translation parameters have been developed and will be described in Sections 2.3.6 and 2.3.7, respectively.

Although as many functions have been used to detect solutions in these Patterson search procedures as there are programs, most rely on some variation of the sum, product and minimum functions (Section 2.3.2.4). The quality of the stored Patterson density representation also varies widely, but it is now common to use 16 or more bits for single density values. Treatment of vector overlap is handled differently by different investigators and the choice will depend on the degree of overlapping (Nordman & Schilling, 1970; Nordman, 1972). General Gaussian multiplicity corrections can be employed to treat coincidental overlap of independent vectors in general positions and overlap which occurs for symmetric peaks in the vicinity of a special position or mirror plane in the Patterson (G. Kamer, S. Ramakumar & P. Argos, unpublished results; Rossmann *et al*., 1972).

2.3.3. Isomorphous replacement difference Pattersons

2.3.3.1. Introduction

One of the initial stages in the application of the isomorphous replacement method is the determination of heavy-atom positions. Indeed, this step of a structure determination can often be the most challenging. Not only may the number of heavy-atom sites be unknown, and have incomplete substitution, but the various isomorphous compounds may also lack isomorphism. To compound these problems, the error in the measurement of the isomorphous difference in structure amplitudes is often comparable to the differences themselves. Clearly, therefore, the ease with which a particular problem can be solved is closely correlated with the quality of the data-measuring procedure.

The isomorphous replacement method was used incidentally by Bragg in the solution of NaCl and KCl. It was later formalized by J. M. Robertson in the analysis of phthalocyanine where the coordination centre could be Pt, Ni and other metals (Robertson, 1935, 1936; Robertson & Woodward, 1937). In this and similar cases, there was no difficulty in finding the heavy-atom positions. Not only were the heavy atoms frequently in special positions, but they dominated the total scattering effect. It was not until Perutz and his colleagues (Green *et al*., 1954; Bragg & Perutz, 1954) applied the technique to the solution of haemoglobin, a protein of 68 000 Da, that it was necessary to consider methods for detecting heavy atoms. The effect of a single heavy atom, even uranium, can only have a very marginal effect on the structure amplitudes of a crystalline macromolecule. Hence, techniques had to be developed which were dependent on the difference of the isomorphous structure amplitudes rather than on the solution of the Patterson of the heavy-atom-derivative compound on its own.

2.3.3.2. Finding heavy atoms with centrosymmetric projections

Phases in a centrosymmetric projection will be 0 or π if the origin is chosen at the centre of symmetry. Hence, the native structure factor, $\mathbf{F}_N$, and the heavy-atom-derivative structure factor, $\mathbf{F}_{NH}$, will be collinear. It follows that the structure amplitude, $|\mathbf{F}_H|$, of the heavy atoms alone in the cell will be given by

$$|\mathbf{F}_H| = |(|\mathbf{F}_{NH}| \pm |\mathbf{F}_N|)| + \varepsilon,$$

where ε is the error on the parenthetic sum or difference. Three different cases may arise (Fig. 2.3.3.1). Since the situation shown in Fig. 2.3.3.1(*c*) is rare, in general

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2. \qquad (2.3.3.1)$$

Thus, a Patterson computed with the square of the differences between the native and derivative structure amplitudes of a centrosymmetric projection will approximate to a Patterson of the heavy atoms alone.

The approximation (2.3.3.1) is valid if the heavy-atom substitution is small enough to make $|\mathbf{F}_H| \ll |\mathbf{F}_{NH}|$ for most reflections, but sufficiently large to make $\varepsilon \ll (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$. It is also

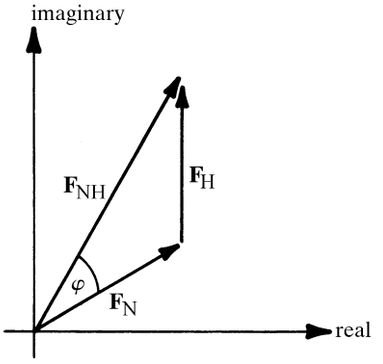


Fig. 2.3.3.2. Vector triangle showing the relationship between $\mathbf{F}_N$, $\mathbf{F}_{NH}$ and $\mathbf{F}_H$, where $\mathbf{F}_{NH} = \mathbf{F}_N + \mathbf{F}_H$.

assumed that the native and heavy-atom-derivative data have been placed on the same relative scale. Hence, the relation (2.3.3.1) should be rewritten as

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - k|\mathbf{F}_N|)^2,$$

where k is an experimentally determined scale factor (see Section 2.3.3.7). Uncertainty in the determination of k will contribute further to ε, albeit in a systematic manner.

Centrosymmetric projections were used extensively for the determination of heavy-atom sites in early work on proteins such as haemoglobin (Green *et al.*, 1954), myoglobin (Bluhm *et al.*, 1958) and lysozyme (Poljak, 1963). However, with the advent of faster data-collecting techniques, low-resolution (*e.g.* a 5 Å limit) three-dimensional data are to be preferred for calculating difference Pattersons. For noncentrosymmetric reflections, the approximation (2.3.3.1) is still valid but less exact (Section 2.3.3.3). However, the larger number of three-dimensional differences compared to projection differences will enhance the signal of the real Patterson peaks relative to the noise. If there are N terms in the Patterson synthesis, then the peak-to-noise ratio will be proportionally $\sqrt{N}$ and $1/\varepsilon$. With the subscripts 2 and 3 representing two- and three-dimensional syntheses, respectively, the latter will be more powerful than the former whenever

$$\frac{\sqrt{N_3}}{\varepsilon_3} > \frac{\sqrt{N_2}}{\varepsilon_2}.$$

Now, as $\varepsilon_3 \simeq \sqrt{2}\varepsilon_2$, it follows that N_3 must be greater than $2N_2$ if the three-dimensional noncentrosymmetric computation is to be more powerful. This condition must almost invariably be true.

2.3.3.3. Finding heavy atoms with three-dimensional methods

A Patterson of a native biomacromolecular structure (coefficients F_N^2) can be considered as being, at least approximately, a vector map of all the light atoms (carbons, nitrogens, oxygens, some sulfurs, and also phosphorus for nucleic acids) other than hydrogen atoms. These interactions will be designated as LL. Similarly, a Patterson of the heavy-atom derivative will contain $HH + HL + LL$ interactions, where H represents the heavy atoms. Thus, a true difference Patterson, with coefficients $F_{NH}^2 - F_N^2$, will contain only the interactions $HH + HL$. In general, the carpet of HL vectors completely dominates the HH vectors except for very small proteins such as insulin (Adams *et al.*, 1969). Therefore, it would be preferable to compute a Patterson containing only HH interactions in order to interpret the map in terms of specific heavy-atom sites.

Blow (1958) and Rossmann (1960) showed that a Patterson with $(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$ coefficients approximated to a Patterson containing only HH vectors. If the phase angle between $\mathbf{F}_N$ and $\mathbf{F}_{NH}$ is φ (Fig. 2.3.3.2), then

$$|\mathbf{F}_H|^2 = |\mathbf{F}_N|^2 + |\mathbf{F}_{NH}|^2 - 2|\mathbf{F}_N||\mathbf{F}_{NH}|\cos\varphi.$$

In general, however, $|\mathbf{F}_H| \ll |\mathbf{F}_N|$. Hence, φ is small and

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2,$$

which is the same relation as (2.3.3.1) for centrosymmetric approximations. Since the direction of $\mathbf{F}_H$ is random compared to $\mathbf{F}_N$, the root-mean-square projected length of $\mathbf{F}_H$ onto $\mathbf{F}_N$ will be $\mathbf{F}_H/\sqrt{2}$. Thus it follows that a better approximation is

$$|\mathbf{F}_H|^2 \simeq \sqrt{2}(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2, \qquad (2.3.3.2)$$

which accounts for the assumption (Section 2.3.3.2) that $\varepsilon_3 = \sqrt{2}\varepsilon_2$. The almost universal method for the initial determination of major heavy-atom sites in an isomorphous derivative utilizes a Patterson with $(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$ coefficients. Approximation (2.3.3.2) is also the basis for the refinement of heavy-atom parameters in a single isomorphous replacement pair (Rossmann, 1960; Cullis *et al.*, 1962; Terwilliger & Eisenberg, 1983).

2.3.3.4. Correlation functions

In the most general case of a triclinic space group, it will be necessary to select an origin arbitrarily, usually coincident with a heavy atom. All other heavy atoms (and subsequently also the macromolecular atoms) will be referred to this reference atom. However, the choice of an origin will be independent in the interpretation of each derivative's difference Patterson. It will then be necessary to correlate the various, arbitrarily chosen, origins. The same problem occurs in space groups lacking symmetry axes perpendicular to the primary rotation axis (*e.g.* $P2_1$, $P6$ *etc.*), although only one coordinate, namely parallel to the unique rotation axis, will require correlation. This problem gave rise to some concern in the 1950s. Bragg (1958), Blow (1958), Perutz (1956), Hoppe (1959) and Bodo *et al.* (1959) developed a variety of techniques, none of which were entirely satisfactory. Rossmann (1960) proposed the $(\mathbf{F}_{NH1} - \mathbf{F}_{NH2})^2$ synthesis and applied it successfully to the heavy-atom determination of horse haemoglobin. This function gives positive peaks ($H1 \cdot H1$) at the end of vectors between the heavy-atom sites in the first compound, positive peaks ($H2 \cdot H2$) between the sites in the second compound, and negative peaks between sites in the first and second compound (Fig. 2.3.3.3). It is thus the negative peaks which provide the necessary correlation. The function is unique in that it is a Patterson containing significant information in both positive and negative peaks. Steinrauf (1963) suggested using the coefficients $(|\mathbf{F}_{NH1}| - |\mathbf{F}_N|) \cdot (|\mathbf{F}_{NH2}| - |\mathbf{F}_N|)$ in order to eliminate the positive $H1 \cdot H1$ and $H2 \cdot H2$ vectors.

Although the problem of correlation was a serious concern in the early structural determination of proteins during the late 1950s and early 1960s, the problem has now been bypassed. Blow & Rossmann (1961) and Kartha (1961) independently showed that it was possible to compute usable phases from a single isomorphous replacement (SIR) derivative. This contradicted the previously accepted notion that it was necessary to have at least two isomorphous derivatives to be able to determine a noncentrosymmetric reflection's phase (Harker, 1956). Hence, currently, the procedure used to correlate origins in different derivatives is to compute SIR phases from the first compound and apply them to a difference electron-density map of the second heavy-atom derivative. Thus, the origin of the second

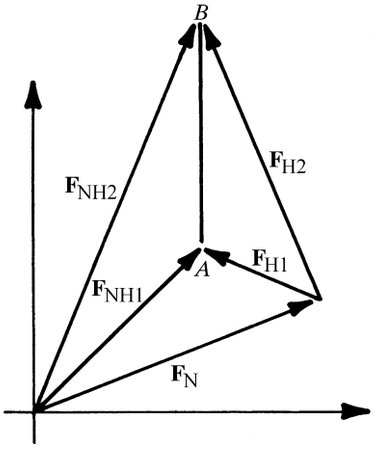


Fig. 2.3.3.3. A Patterson with coefficients $(\mathbf{F}_{NH1} - \mathbf{F}_{NH2})^2$ will be equivalent to a Patterson whose coefficients are $(AB)^2$. However, $AB = -\mathbf{F}_{H1} + \mathbf{F}_{H2}$. Thus, a Patterson with $(AB)^2$ coefficients is equivalent to having negative atomic substitutions in compound 1 and positive substitutions in compound 2, or *vice versa*. Therefore, the Patterson will contain positive peaks for vectors of the type $H1 \cdot H1$ and $H2 \cdot H2$, but negative vector peaks for vectors of type $H1 \cdot H2$.

derivative will be referred to the arbitrarily chosen origin of the first compound. More important, however, the interpretation of such a 'feedback' difference Fourier is easier than that of a difference Patterson. Hence, once one heavy-atom derivative has been solved for its heavy-atom sites, the solution of other derivatives is almost assured. This concept is examined more closely in the following section.

2.3.3.5. Interpretation of isomorphous difference Pattersons

Difference Pattersons have usually been manually interpreted in terms of point atoms. In more complex situations, such as crystalline viruses, a systematic approach may be necessary to analyse the Patterson. That is especially true when the structure contains noncrystallographic symmetry (Argos & Rossmann, 1976). Such methods are in principle dependent on the comparison of the observed Patterson, $P_1(\mathbf{x})$, with a calculated Patterson, $P_2(\mathbf{x})$. A criterion, C_P, based on the sum of the Patterson densities at all test vectors within the unit-cell volume V, would be

$$C_P = \int_V P_1(\mathbf{x}) \cdot P_2(\mathbf{x})\, d\mathbf{x}.$$

C_P can be evaluated for all reasonable heavy-atom distributions. Each different set of trial sites corresponds to a different P_2 Patterson. It is then easily shown that

$$C_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 E_{\mathbf{h}}^2,$$

where the sum is taken over all $\mathbf{h}$ reflections in reciprocal space, $\Delta_{\mathbf{h}}^2$ are the observed differences and $E_{\mathbf{h}}$ are the structure factors of the trial point Patterson. (The symbol E is used here because of its close relation to normalized structure factors.)

Let there be n noncrystallographic asymmetric units within the crystallographic asymmetric unit and m crystallographic asymmetric units within the crystal unit cell. Then there are L symmetry-related heavy-atom sites where $L = nm$. Let the scattering contribution of the ith site have a_i and b_i real and imaginary structure-factor components with respect to an arbitrary origin. Hence, for reflection $\mathbf{h}$

$$E_{\mathbf{h}}^2 \left(\sum_L a_{\mathbf{h}i}\right)^2 + \left(\sum_L b_{\mathbf{h}i}\right)^2 = L + \sum_{i\neq j}^{N}\sum^{N} (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}).$$

Therefore,

$$C_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[L + 2 \sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right].$$

But $\sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2$ must be independent of the number, L, of heavy-atom sites per cell. Thus the criterion can be rewritten as

$$C'_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[\sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right]. \quad (2.3.3.3)$$

More generally, if some sites have already been tentatively determined, and if these sites give rise to the structure-factor components $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$, then

$$E_{\mathbf{h}}^2 = \left(A_{\mathbf{h}} + \sum_N a_{\mathbf{h}i}\right)^2 + \left(B_{\mathbf{h}} + \sum_N b_{\mathbf{h}i}\right)^2. \quad (2.3.3.4)$$

Following the same procedure as above, it follows that

$$C'_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[(A_{\mathbf{h}}a_{\mathbf{h}} + B_{\mathbf{h}}b_{\mathbf{h}}) + \sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right], \quad (2.3.3.5)$$

where $a_{\mathbf{h}} = \sum_{i=1}^{L} a_{\mathbf{h}i}$ and $b_{\mathbf{h}} = \sum_{i=1}^{L} b_{\mathbf{h}i}$.

Expression (2.3.3.5) will now be compared with the 'feedback' method (Dickerson *et al.*, 1967, 1968) of verifying heavy-atom sites using SIR phasing. Inspection of Fig. 2.3.3.4 shows that the native phase, α, will be determined as $\alpha = \varphi + \pi$ (φ is the structure-factor phase corresponding to the presumed heavy-atom positions) when $|\mathbf{F}_N| > |\mathbf{F}_H|$ and $\alpha = \varphi$ when $|\mathbf{F}_N| \ll |\mathbf{F}_H|$. Thus, an SIR difference electron density, $\Delta\rho(\mathbf{x})$, can be synthesized by the Fourier summation

$$\begin{aligned}\Delta\rho(\mathbf{x}) =& \frac{1}{V}\sum m(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}}) \\ & \text{from terms with } \Delta_{\mathbf{h}} = |\mathbf{F}_{NH}| - |\mathbf{F}_N| > 0 \\ & + \frac{1}{V}\sum m(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}} - \pi) \\ & \text{from terms with } \Delta_{\mathbf{h}} < 0 \\ =& \frac{1}{V}\sum m|\Delta_{\mathbf{h}}|\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}}),\end{aligned}$$

where m is a figure of merit of the phase reliability (Blow & Crick, 1959; Dickerson *et al.*, 1961). Now,

$$\mathbf{F}_{\mathbf{h}} = A_{\mathbf{h}} + iB_{\mathbf{h}} = F_H \cos\varphi_{\mathbf{h}} + iF_H \sin\varphi_{\mathbf{h}},$$

where $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ are the real and imaginary components of the presumed heavy-atom sites. Therefore,

$$\Delta\rho(\mathbf{x}) = \frac{1}{V}\sum \frac{m|\Delta_{\mathbf{h}}|}{|\mathbf{F}_H|}(A_{\mathbf{h}}\cos 2\pi\mathbf{h}\cdot\mathbf{x} + B_{\mathbf{h}}\sin 2\pi\mathbf{h}\cdot\mathbf{x}).$$

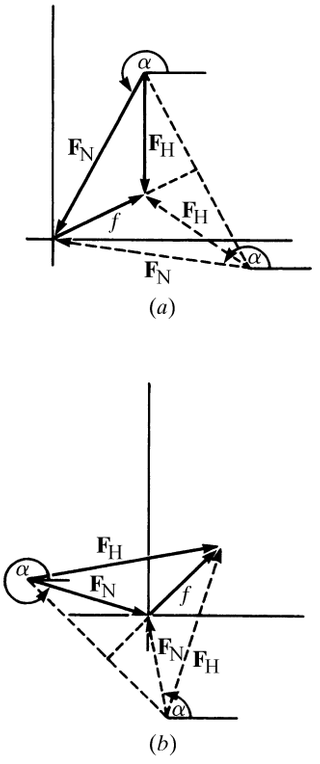


Fig. 2.3.3.4. The phase α of the native compound (structure factor $\mathbf{F}_N$) is determined either as being equal to, or 180° out of phase with, the presumed heavy-atom contribution when only a single isomorphous compound is available. In (*a*) is shown the case when $|\mathbf{F}_N| > |\mathbf{F}_{NH}|$ and $\alpha \simeq \varphi + \pi$. In (*b*) is shown the case when $|\mathbf{F}_N| < |\mathbf{F}_{NH}|$ and $\alpha = \varphi$, where φ is the phase of the heavy-atom structure factor $\mathbf{F}_H$.

If this SIR difference electron-density map shows significant peaks at sites related by noncrystallographic symmetry, then those sites will be at the position of a further set of heavy atoms. Hence, a suitable criterion for finding heavy-atom sites is

$$C_{\rm SIR} = \sum_{j=1}^{n} \Delta\rho(\mathbf{x}_j),$$

or by substitution

$$C_{\rm SIR} = \sum_{j=1}^{n} \frac{1}{V} \sum_{\mathbf{h}} \frac{m|\Delta_{\mathbf{h}}|}{|\mathbf{F}_H|} (A_{\mathbf{h}} \cos 2\pi\mathbf{h}\cdot\mathbf{x}_j + B_{\mathbf{h}} \sin 2\pi\mathbf{h}\cdot\mathbf{x}_j).$$

But

$$a_{\mathbf{h}} = \sum_{j=1}^{n} \cos 2\pi\mathbf{h}\cdot\mathbf{x}_j \quad \text{and} \quad b_{\mathbf{h}} = \sum_{j=1}^{n} \sin 2\pi\mathbf{h}\cdot\mathbf{x}_j.$$

Therefore,

$$C_{\rm SIR} = \frac{1}{V} \sum_{\mathbf{h}} \frac{m|\Delta_{\mathbf{h}}|}{|\mathbf{F}_H|} (A_{\mathbf{h}} a_{\mathbf{h}} + B_{\mathbf{h}} b_{\mathbf{h}}). \qquad (2.3.3.6)$$

This expression is similar to (2.3.3.5) derived by consideration of a Patterson search. It differs from (2.3.3.5) in two respects: the Fourier coefficients are different and expression (2.3.3.6) is lacking a second term. Now the figure of merit m will be small whenever $|\mathbf{F}_H|$ is small as the SIR phase cannot be determined well under those conditions. Hence, effectively, the coefficients are a function of $|\Delta_{\mathbf{h}}|$, and the coefficients of the functions

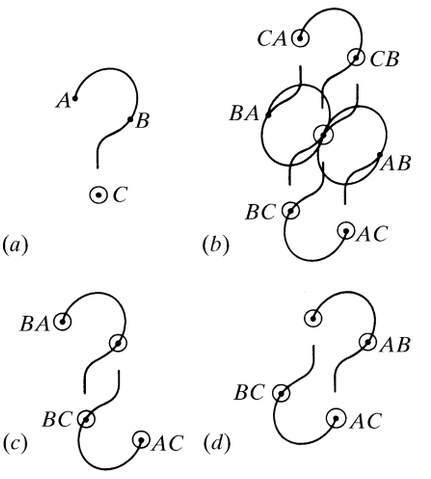


Fig. 2.3.3.5. Let (*a*) be the original structure which contains three heavy atoms *ABC* in a noncentrosymmetric configuration. Then a Fourier summation, with $(F_{NH}^2 - F_N^2)$ coefficients, will give the Patterson shown in (*b*). Displacement of the Patterson by the vector *BC* and selecting the common patterns yields (*c*). Similarly, displacement by *AC* gives (*d*). Finally, superposition of (*c*) on (*d*) gives the original figure or its enantiomorph. This series of steps demonstrates that, in principle, complete structural information is contained in an SIR derivative.

(2.3.3.5) and (2.3.3.6) are indeed rather similar. The second term in (2.3.3.5) relates to the use of the search atoms in phasing and could be included in (2.3.3.6), provided the actual feedback sites in each of the n electron-density functions tested by $C_{\rm SIR}$ are omitted in turn. Thus, a systematic Patterson search and an SIR difference Fourier search are very similar in character and power.

2.3.3.6. Direct structure determination from difference Pattersons

The difference Patterson computed with coefficients $F_{HN}^2 - F_N^2$ contains information on the heavy atoms (*HH* vectors) and the macromolecular structure (*HL* vectors) (Section 2.3.3.3). If the scaling between the $|\mathbf{F}_{HN}|$ and $|\mathbf{F}_N|$ data sets is not perfect there will also be noise. Rossmann (1961*b*) was partially successful in determining the low-resolution horse haemoglobin structure by using a series of superpositions based on the known heavy-atom sites. Nevertheless, Patterson superposition methods have not been used for the structure determination of proteins owing to the successful error treatment of the isomorphous replacement method in reciprocal space. However, it is of some interest here for it gives an alternative insight into SIR phasing.

The deconvolution of an arbitrary molecule, represented as '?', from an $(F_{HN}^2 - F_N^2)$ Patterson, is demonstrated in Fig. 2.3.3.5. The original structure is shown in Fig. 2.3.3.5(*a*) and the corresponding Patterson in Fig. 2.3.3.5(*b*). Superposition with respect to one of the heavy-atom sites is shown in Fig. 2.3.3.5(*c*) and the other in Fig. 2.3.3.5(*d*). Both Figs. 2.3.3.5(*c*) and (*d*) contain a centre of symmetry because the use of only a single *HH* vector implies a centre of symmetry half way between the two sites. The centre is broken on combining information from all three sites (which together lack a centre of symmetry) by superimposing Figs. 2.3.3.5(*c*) and (*d*) to obtain either the original structure (Fig. 2.3.3.5*a*) or its enantiomorph. Thus it is clear, in principle, that there is sufficient information in a single isomorphous derivative data set, when used in conjunction with a native data set, to solve a structure completely. However, the procedure shown in Fig. 2.3.3.5 does not consider the accumulation of error in the selection of individual images when these intersect with another image. In this sense the reciprocal-space isomorphous replacement technique has greater elegance and provides more insight, whereas the alternative view given by the Patterson method was the original stimulus for the discovery of the SIR phasing technique (Blow & Rossmann, 1961).

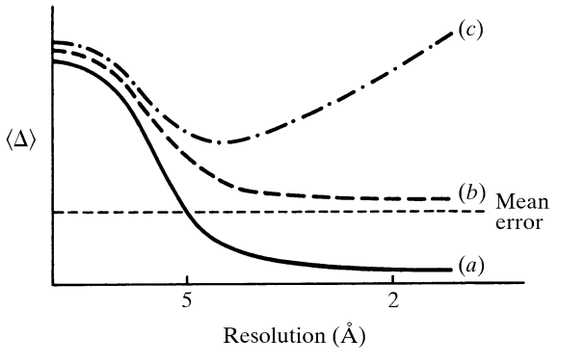


Fig. 2.3.3.6. A plot of mean isomorphous differences as a function of resolution. (*a*) The theoretical size of mean differences following roughly a Gaussian distribution. (*b*) The observed size of differences for a good isomorphous derivative where the smaller higher-order differences have been largely masked by the error of measurement. (*c*) Observed differences where 'lack of isomorphism' dominates beyond approximately 5 Å resolution.

Other Patterson functions for the deconvolution of SIR data have been proposed by Ramachandran & Raman (1959), as well as others. The principles are similar but the coefficients of the functions are optimized to emphasize various aspects of the signal representing the molecular structure.

2.3.3.7. Isomorphism and size of the heavy-atom substitution

It is insufficient to discuss Patterson techniques for locating heavy-atom substitutions without also considering errors of all kinds. First, it must be recognized that most heavy-atom labels are not a single atom but a small compound containing one or more heavy atoms. The compound itself will displace water or ions and locally alter the conformation of the protein or nucleic acid. Hence, a simple Gaussian approximation will suffice to represent individual heavy-atom scatterers responsible for the difference between native and heavy-atom derivatives. Furthermore, the heavy-atom compound often introduces small global structural changes which can be detected only at higher resolution. These problems were considered with some rigour by Crick & Magdoff (1956). In general, lack of isomorphism is exhibited by an increase in the size of the isomorphous differences with increasing resolution (Fig. 2.3.3.6).

Crick & Magdoff (1956) also derived the approximate expression

$$\sqrt{\frac{2N_{\rm H}}{N_{\rm P}}}\cdot\frac{f_{\rm H}}{f_{\rm P}}$$

to estimate the r.m.s. fractional change in intensity as a function of heavy-atom substitution. Here, $N_{\rm H}$ represents the number of heavy atoms attached to a protein (or other large molecule) which contains $N_{\rm P}$ light atoms. $f_{\rm H}$ and $f_{\rm P}$ are the scattering powers of the average heavy and protein atom, respectively. This function was tabulated by Eisenberg (1970) as a function of molecular weight (proportional to $N_{\rm P}$). For instance, for a single, fully substituted, Hg atom the formula predicts an r.m.s. intensity change of around 25% in a molecule of 100 000 Da. However, the error of measurement of a reflection intensity is likely to be arround 10% of I, implying perhaps an error of around 14% of I on a difference measurement. Thus, the isomorphous replacement difference measurement for almost half the reflections will be buried in error for this case.

Scaling of the different heavy-atom-derivative data sets onto a common relative scale is clearly important if error is to be reduced. Blundell & Johnson (1976, pp. 333–336) give a careful discussion of this subject. Suffice it to say here only that a linear scale factor is seldom acceptable as the heavy-atom-derivative crystals frequently suffer from greater disorder than the native crystals. The heavy-atom derivative should, in general, have a slightly larger mean value for the structure factors on account of the additional heavy atoms (Green *et al.*, 1954). The usual effect is to make $\sum|\mathbf{F}_{\rm NH}|^2/\sum|\mathbf{F}_{\rm N}|^2 \simeq 1.05$ (Phillips, 1966).

As the amount of heavy atom is usually unknown in a yet unsolved heavy-atom derivative, it is usual practice either to apply a scale factor of the form $k\exp[-B(\sin\theta/\lambda)^2]$ or, more generally, to use local scaling (Matthews & Czerwinski, 1975). The latter has the advantage of not making any assumption about the physical nature of the relative intensity decay with resolution.

2.3.4. Anomalous dispersion

2.3.4.1. Introduction

The physical basis for anomalous dispersion has been well reviewed by Ramaseshan & Abrahams (1975), James (1965), Cromer (1974) and Bijvoet (1954). As the wavelength of radiation approaches the absorption edge of a particular element, then an atom will disperse X-rays in a manner that can be defined by the complex scattering factor

$$f_0 + \Delta f' + i\Delta f'',$$

where f_0 is the scattering factor of the atom without the anomalous absorption and rescattering effect, $\Delta f'$ is the real correction term (usually negative), and $\Delta f''$ is the imaginary component. The real term $f_0 + \Delta f'$ is often written as f', so that the total scattering factor will be $f' + i\Delta f''$. Values of $\Delta f'$ and $\Delta f''$ are tabulated in *IT* IV (Cromer, 1974), although their precise values are dependent on the environment of the anomalous scatterer. Unlike f_0, $\Delta f'$ and $\Delta f''$ are almost independent of scattering angle as they are caused by absorption of energy in the innermost electron shells. Thus, the anomalous effect resembles scattering from a point atom.

The structure factor of index **h** can now be written as

$$\mathbf{F}_{\mathbf{h}} = \sum_{j=1}^{N} f_j' \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j) + i\sum_{j=1}^{N}\Delta f_j'' \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j). \qquad (2.3.4.1)$$

(Note that the only significant contributions to the second term are from those atoms that have a measurable anomalous effect at the chosen wavelength.)

Let us now write the first term as $A + iB$ and the second as $a + ib$. Then, from (2.3.4.1),

$$\mathbf{F} = (A + iB) + i(a + ib) = (A - b) + i(B + a). \qquad (2.3.4.2)$$

Therefore,

$$|\mathbf{F}_{\mathbf{h}}|^2 = (A - b)^2 + (B + a)^2$$

and similarly

$$|\mathbf{F}_{\bar{\mathbf{h}}}|^2 = (A + b)^2 + (-B + a)^2,$$

demonstrating that Friedel's law breaks down in the presence of anomalous dispersion. However, it is only for noncentrosymmetric reflections that $|\mathbf{F}_{\mathbf{h}}| \neq |\mathbf{F}_{\bar{\mathbf{h}}}|$.

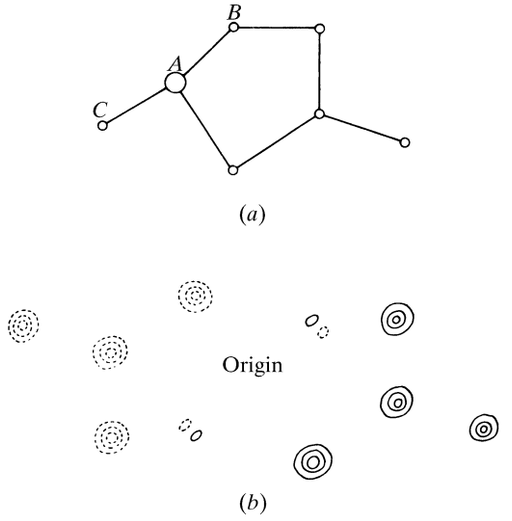


Fig. 2.3.4.1. (*a*) A model structure with an anomalous scatterer at *A*. (*b*) The corresponding $P_s(\mathbf{u})$ function showing positive peaks (full lines) and negative peaks (dashed lines). [Reprinted with permission from Woolfson (1970, p. 293).]

Now,

$$\rho(\mathbf{x}) = \frac{1}{V}\sum_{\mathbf{h}}^{\text{sphere}} \mathbf{F}_{\mathbf{h}} \exp(2\pi i \mathbf{h}\cdot\mathbf{x}).$$

Hence, by using (2.3.4.2) and simplifying,

$$\rho(\mathbf{x}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} [(A\cos 2\pi\mathbf{h}\cdot\mathbf{x} - B\sin 2\pi\mathbf{h}\cdot\mathbf{x}) + i(a\cos 2\pi\mathbf{h}\cdot\mathbf{x} - b\sin 2\pi\mathbf{h}\cdot\mathbf{x})]. \quad (2.3.4.3)$$

The first term in (2.3.4.3) is the usual real Fourier expression for electron density, while the second term is an imaginary component due to the anomalous scattering of a few atoms in the cell.

2.3.4.2. The $P_s(\mathbf{u})$ function

Expression (2.3.4.3) gives the complex electron density expression in the presence of anomalous scatterers. A variety of Patterson-type functions can be derived from (2.3.4.3) for the determination of a structure. For instance, the $P_s(\mathbf{u})$ function gives vectors between the anomalous atoms and the 'normal' atoms.

From (2.3.4.1) it is easy to show that

$$\begin{aligned}\mathbf{F}_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}^* &= |\mathbf{F}_{\mathbf{h}}|^2 \\ &= \sum_{i,j}(f_i' f_j' + f_i'' f_j'')\cos 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j) \\ &\quad + \sum_{i,j}(f_i' f_j'' - f_i'' f_j')\sin 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j).\end{aligned}$$

Therefore,

$$|\mathbf{F}_{\mathbf{h}}|^2 + |\mathbf{F}_{\bar{\mathbf{h}}}|^2 = 2\sum_{i,j}(f_i' f_j' + f_i'' f_j'')\cos 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j)$$

and

$$|\mathbf{F}_{\mathbf{h}}|^2 - |\mathbf{F}_{\bar{\mathbf{h}}}|^2 = 2\sum_{i,j}(f_i' f_j'' - f_i'' f_j')\sin 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j).$$

Let us now consider the significance of a Patterson in the presence of anomalous dispersion. The normal Patterson definition is given by

$$\begin{aligned}P(\mathbf{u}) &= \int_V \rho^*(\mathbf{x})\rho(\mathbf{x}+\mathbf{u})\,\mathrm{d}\mathbf{x} \\ &= \frac{1}{V^2}\sum_{\mathbf{h}}^{\text{sphere}} |\mathbf{F}_{\mathbf{h}}|^2 \exp(-2\pi i\mathbf{h}\cdot\mathbf{u}) \\ &\equiv P_c(\mathbf{u}) - iP_s(\mathbf{u}),\end{aligned}$$

where

$$P_c(\mathbf{u}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} (|\mathbf{F}_{\mathbf{h}}|^2 + |\mathbf{F}_{\bar{\mathbf{h}}}|^2)\cos 2\pi\mathbf{h}\cdot\mathbf{u}$$

and

$$P_s(\mathbf{u}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} (|\mathbf{F}_{\mathbf{h}}|^2 - |\mathbf{F}_{\bar{\mathbf{h}}}|^2)\sin 2\pi\mathbf{h}\cdot\mathbf{u}.$$

The $P_c(\mathbf{u})$ component is essentially the normal Patterson, in which the peak heights have been very slightly modified by the anomalous-scattering effect. That is, the peaks of $P_c(\mathbf{u})$ are proportional in height to $(f_i' f_j' + f_i'' f_j'')$.

The $P_s(\mathbf{u})$ component is more interesting. It represents vectors between the normal atoms in the unit cell and the anomalous scatterers, proportional in height to $(f_i' f_j'' - f_i'' f_j')$ (Okaya *et al.*, 1955). This function is antisymmetric with respect to the change of the direction of the diffraction vector. An illustration of the function is given in Fig. 2.3.4.1. In a unit cell containing N atoms, n of which are anomalous scatterers, the $P_s(\mathbf{u})$ function contains only $n(N-n)$ positive peaks and an equal number of negative peaks related to the former by anticentrosymmetry. The analysis of a $P_s(\mathbf{u})$ synthesis presents problems somewhat similar to those posed by a normal Patterson. The procedure has been used only rarely [*cf.* Moncrief & Lipscomb (1966) and Pepinsky *et al.* (1957)], probably because alternative procedures are available for small compounds and the solution of $P_s(\mathbf{u})$ is too complex for large biological molecules.

2.3.4.3. The position of anomalous scatterers

Anomalous scatterers can be used as an aid to phasing, when their positions are known. But the anomalous-dispersion differences (Bijvoet differences) can also be used to determine or confirm the heavy atoms which scatter anomalously (Rossmann, 1961*a*). Furthermore, the use of anomalous-dispersion information obviates the lack of isomorphism but, on the other hand, the differences are normally far smaller than those produced by a heavy-atom isomorphous replacement.

Consider a structure of many light atoms giving rise to the structure factor $\mathbf{F}_{\mathbf{h}}(N)$. In addition, it contains a few heavy atoms which have a significant anomalous-scattering effect. The non-anomalous component will be $\mathbf{F}_{\mathbf{h}}(H)$ and the anomalous component is $\mathbf{F}_{\mathbf{h}}''(H) = i(\Delta f''/f')\mathbf{F}_{\mathbf{h}}(H)$ (Fig. 2.3.4.2*a*). The total structure factor will be $\mathbf{F}_{\mathbf{h}}$. The Friedel opposite is constructed appropriately (Fig. 2.3.4.2*a*). Now reflect the Friedel opposite construction across the real axis of the Argand diagram (Fig.

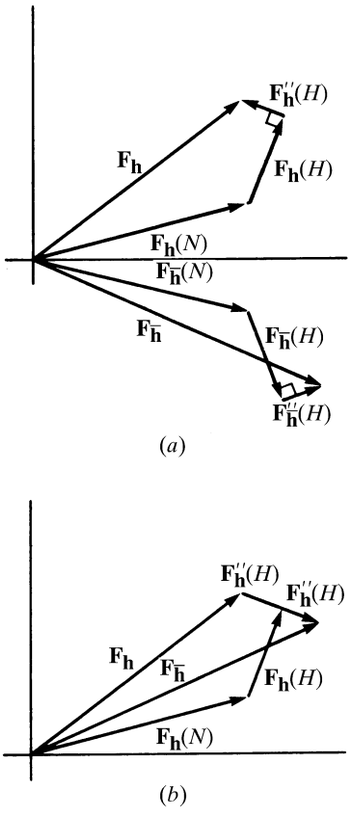


Fig. 2.3.4.2. Anomalous-dispersion effect for a molecule whose light atoms contribute $\mathbf{F}_{\mathbf{h}}(N)$ and heavy atom $\mathbf{F}_{\mathbf{h}}(H)$ with a small anomalous component of $\mathbf{F}_{\mathbf{h}}''(H)$, 90° ahead of the non-anomalous $\mathbf{F}_{\mathbf{h}}(H)$ component. In (*a*) is seen the construction for $\mathbf{F}_{\mathbf{h}}$ and $\mathbf{F}_{\bar{\mathbf{h}}}$. In (*b*) $\mathbf{F}_{\bar{\mathbf{h}}}$ has been reflected across the real axis.

2.3.4.2*b*). Let the difference in phase between $\mathbf{F}_{\mathbf{h}}$ and $\mathbf{F}_{\bar{\mathbf{h}}}$ be φ. Thus

$$4|\mathbf{F}_{\mathbf{h}}''(H)|^2 = |\mathbf{F}_{\mathbf{h}}|^2 + |\mathbf{F}_{\bar{\mathbf{h}}}|^2 - 2|\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\bar{\mathbf{h}}}|\cos\varphi,$$

but since φ is very small

$$|\mathbf{F}_{\mathbf{h}}''(H)|^2 \simeq \tfrac{1}{4}(|\mathbf{F}_{\mathbf{h}}| - |\mathbf{F}_{\bar{\mathbf{h}}}|)^2.$$

Hence, a Patterson with coefficients $(|\mathbf{F}_{\mathbf{h}}| - |\mathbf{F}_{\bar{\mathbf{h}}}|)^2$ will be equivalent to a Patterson with coefficients $|\mathbf{F}_{\mathbf{h}}''(H)|^2$ which is proportional to $|\mathbf{F}_{\mathbf{h}}(H)|^2$. Such a Patterson (Rossmann, 1961*a*) will have vectors between all anomalous scatterers with heights proportional to the number of anomalous electrons $\Delta f''$. This 'anomalous dispersion' Patterson function has been used to find anomalous scatterers such as iron (Smith *et al.*, 1983; Strahs & Kraut, 1968) and sulfur atoms (Hendrickson & Teeter, 1981). The anomalous signal from Se atoms in selenomethionine-substituted proteins has been found to be extremely powerful and is now routinely used for protein structure determinations (Hendrickson, 1991). Anomalous signals from halide ions or xenon atoms have also been used to solve protein structures (Dauter *et al.*, 2000; Nagem *et al.*, 2003; Schiltz *et al.*, 2003). The anomalous signal from sulfur atoms, though very small (Hendrickson & Teeter, 1981), has recently been applied successfully to solve several protein structures (Debreczeni *et al.*, 2003; Ramagopal *et al.*, 2003; Yang *et al.*, 2003).

It is then apparent that a Patterson with coefficients

$$\Delta F_{\mathrm{ANO}}^2 = (|\mathbf{F}_{\mathbf{h}}| - |\mathbf{F}_{\bar{\mathbf{h}}}|)^2$$

(Rossmann, 1961*a*), as well as a Patterson with coefficients

$$\Delta F_{\mathrm{ISO}}^2 = (|\mathbf{F}_{\mathrm{NH}}| - |\mathbf{F}_{\mathrm{H}}|)^2$$

(Rossmann, 1960; Blow, 1958), represent Pattersons of the heavy atoms. The $\Delta F_{\mathrm{ANO}}^2$ Patterson suffers from errors which may be larger than the size of the Bijvoet differences, while the $\Delta F_{\mathrm{ISO}}^2$ Patterson may suffer from partial lack of isomorphism. Hence, Kartha & Parthasarathy (1965) have suggested the use of the sum of these two Pattersons, which would then have coefficients $(\Delta F_{\mathrm{ANO}}^2 + \Delta F_{\mathrm{ISO}}^2)$.

However, given both SIR and anomalous-dispersion data, it is possible to make an accurate estimate of the $|\mathbf{F}_{\mathrm{H}}|^2$ magnitudes for use in a Patterson calculation [Blundell & Johnson (1976, p. 340), Matthews (1966), Singh & Ramaseshan (1966)]. In essence, the Harker phase diagram can be constructed out of three circles: the native amplitude and each of the two isomorphous Bijvoet differences, giving three circles in all (Blow & Rossmann, 1961) which should intersect at a single point thus resolving the ambiguity in the SIR data and the anomalous-dispersion data. Furthermore, the phase ambiguities are orthogonal; thus the two data sets are cooperative. It can be shown (Matthews, 1966; North, 1965) that

$$F_{\mathrm{N}}^2 = F_{\mathrm{NH}}^2 + F_{\mathrm{N}}^2 \mp \frac{2}{k}(16k^2F_{\mathrm{P}}^2F_{\mathrm{H}}^2 - \Delta I^2)^{1/2},$$

where $\Delta I = F_{\mathrm{NH}}^{+\,2} - F_{\mathrm{NH}}^{-\,2}$ and $k = \Delta f''/f'$. The sign in the third-term expression is $-$ when $|(\alpha_{\mathrm{NH}} - \alpha_{\mathrm{H}})| < \pi/2$ or $+$ otherwise. Since, in general, $|\mathbf{F}_{\mathrm{H}}|$ is small compared to $|\mathbf{F}_{\mathrm{N}}|$, it is reasonable to assume that the sign above is usually negative. Hence, the heavy-atom lower estimate (HLE) is usually written as

$$F_{\mathrm{HLE}}^2 = F_{\mathrm{NH}}^2 + F_{\mathrm{H}}^2 - \frac{2}{k}(16k^2F_{\mathrm{P}}^2F_{\mathrm{H}}^2 - \Delta I^2)^{1/2},$$

which is an expression frequently used to derive Patterson coefficients useful in the determination of heavy-atom positions when both SIR and anomalous-dispersion data are available.

2.3.4.4. Computer programs for automated location of atomic positions from Patterson maps

Several programs are currently used for automated systematic interpretation of (difference) Patterson maps to locate the positions of heavy atoms and/or anomalous scatterers from isomorphous replacement and anomalous-dispersion data (Weeks *et al.*, 2003). These include *Solve* (Terwilliger & Berendzen, 1999), *CNS* (Brünger *et al.*, 1998), *CCP4* (Collaborative Computational Project, Number 4, 1994) and *Patsol* (Tong & Rossmann, 1993). In these programs, sets of trial atomic positions (seeds) are produced based on one- and two-atom solutions to the Patterson map (see Section 2.3.2.5) (Grosse-Kunstleve & Brunger, 1999; Terwilliger *et al.*, 1987; Tong & Rossmann, 1993). Information from a translation search with a single atom can also be used in this process (Grosse-Kunstleve & Brunger, 1999). Scoring functions have been devised to identify the likely correct solutions, based on agreements with the Patterson map or the observed isomorphous or anomalous differences, as well as the quality of the resulting electron-density map (Terwilliger, 2003*b*; Terwilliger & Berendzen, 1999). The power of modern computers allows the rapid screening of a large collection of trial structures, and the correct solution is found automatically in many cases, even when there is a large number of atomic positions (Weeks *et al.*, 2003).

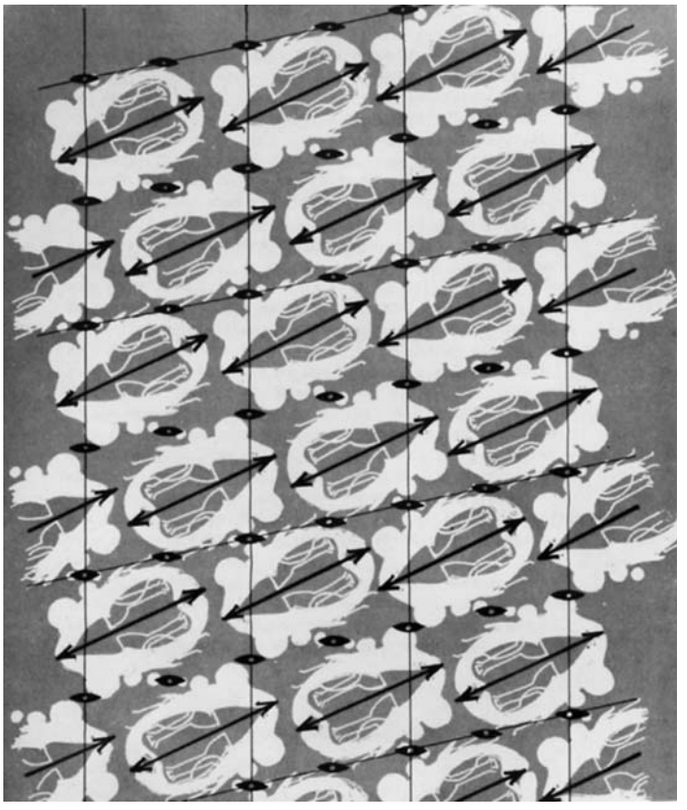


Fig. 2.3.5.1. The two-dimensional periodic design shows crystallographic twofold axes perpendicular to the page and local noncrystallographic rotation axes in the plane of the paper (design by Audrey Rossmann). [Reprinted with permission from Rossmann (1972, p. 8).]

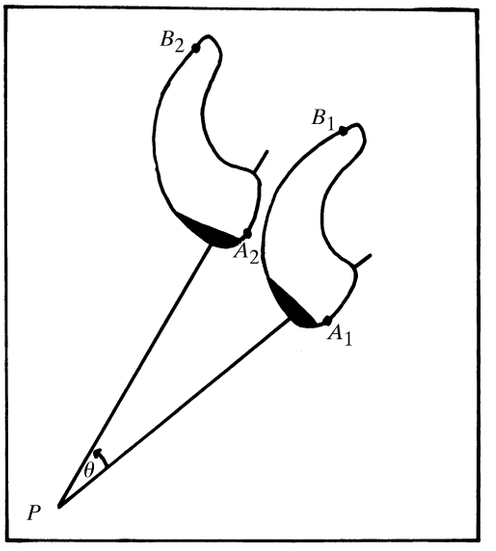


Fig. 2.3.5.2. The objects A_1B_1 and A_2B_2 are related by an improper rotation θ, since it is necessary to consider the sense of rotation to achieve superposition of the two objects. [Reprinted with permission from Rossmann (1972, p. 9).]

2.3.5. Noncrystallographic symmetry

2.3.5.1. Definitions

The interpretation of Pattersons can be helped by using various types of chemical or physical information. An obvious example is the knowledge that one or two heavy atoms per crystallographic asymmetric unit are present. Another example is the exploitation of a rigid chemical framework in a portion of a molecule (Nordman & Nakatsu, 1963; Burnett & Rossmann, 1971). One extremely useful constraint on the interpretation of Pattersons is noncrystallographic symmetry. Indeed, the structural solution of large biological assemblies such as viruses is only possible because of the natural occurrence of this phenomenon. The term 'molecular replacement' is used for methods that utilize noncrystallographic symmetry in the solution of structures [for earlier reviews see Rossmann (1972); Argos & Rossmann (1980); and Rossmann (1990, 2001)]. These methods, which are only partially dependent on Patterson concepts, are discussed in Sections 2.3.6–2.3.8.

Crystallographic symmetry applies to the whole of the three-dimensional crystal lattice. Hence, the symmetry must be expressed both in the lattice and in the repeating pattern within the lattice. In contrast, noncrystallographic symmetry is valid only within a limited volume about the noncrystallographic symmetry element. For instance, the noncrystallographic twofold axes in the plane of the paper of Fig. 2.3.5.1 are true only in the immediate vicinity of each local dyad. In contrast, the crystallographic twofold axes perpendicular to the plane of the paper (Fig. 2.3.5.1) apply to every point within the lattice. Two types of noncrystallographic symmetry can be recognized: proper and improper rotations. A proper symmetry element is independent of the sense of rotation, as, for example, a fivefold axis in an icosahedral virus; a rotation either left or right by one-fifth of a revolution will leave all parts of a given icosahedral shell (but not the whole crystal) in equivalent positions. Proper noncrystallographic symmetry can also be recognized by the existence of a closed point group within a defined volume of the lattice. Improper rotation axes are found when two molecules are arbitrarily oriented relative to each other in the same asymmetric unit or when they occur in two entirely different crystal lattices. For instance, in Fig. 2.3.5.2, the object A_1B_1 can be rotated by $+\theta$ about the axis at P to orient it identically with A_2B_2. However, the two objects will not be coincident after a rotation of A_1B_1 by $-\theta$ or of A_2B_2 by $+\theta$. The envelope around each noncrystallographic object must be known in order to define an improper rotation. In contrast, only the volume about the closed point group need be defined for proper noncrystallographic operations. Hence, the boundaries of the repeating unit need not correspond to chemically covalently linked units in the presence of proper rotations.

Translational components of noncrystallographic rotation elements are said to be 'precise' in a direction parallel to the axis and 'imprecise' perpendicular to the axis (Rossmann *et al.*, 1964). The position, but not direction, of a rotation axis is arbitrary. However, a convenient choice is one that leaves the translation perpendicular to the axis at zero after rotation (Fig. 2.3.5.3).

Noncrystallographic symmetry has been used as a tool in structural analysis primarily in the study of biological molecules. This is due to the propensity of proteins to form aggregates with closed point groups, as, for instance, viruses with 532 symmetry. At best, only part of such a point group can be incorporated into the crystal lattice. Since biological materials cannot contain inversion elements, all studies of noncrystallographic symmetries have been limited to rotational axes. Reflection planes and inversion centres could also be considered in the application of molecular replacement to nonbiological materials.

In this chapter, the relationship

$$\mathbf{x}' = [\boldsymbol{C}]\mathbf{x} + \mathbf{d}$$

will be used to describe noncrystallographic symmetry, where $\mathbf{x}$ and $\mathbf{x}'$ are position vectors, expressed as fractional coordinates, with respect to the crystallographic origin, $[\boldsymbol{C}]$ is a rotation matrix, and $\mathbf{d}$ is a translation vector. Crystallographic symmetry will be described as

$$\mathbf{x}' = [\boldsymbol{T}]\mathbf{x} + \mathbf{t},$$

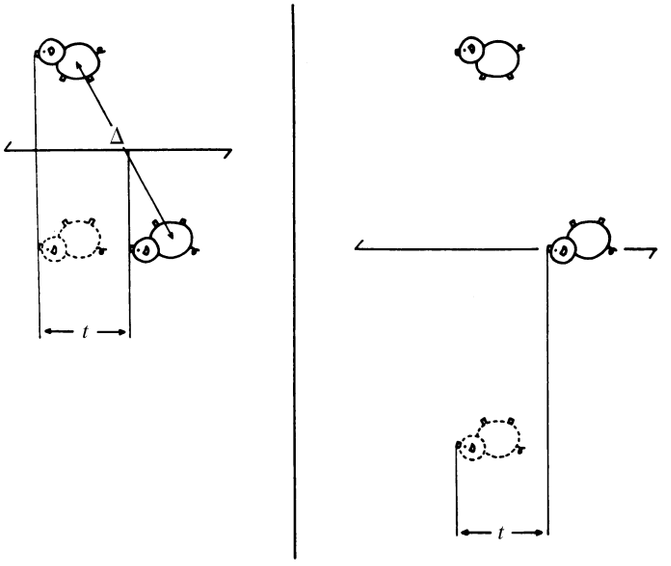


Fig. 2.3.5.3. The position of the twofold rotation axis which relates the two piglets is completely arbitrary. The diagram on the left shows the situation when the translation is parallel to the rotation axis. The diagram on the right has an additional component of translation perpendicular to the rotation axis, but the component parallel to the axis remains unchanged. [Reprinted from Rossmann *et al.* (1964).]

where $[\boldsymbol{T}]$ and $\mathbf{t}$ are the crystallographic rotation matrix and translation vector, respectively. The noncrystallographic asymmetric unit will be defined as having n copies within the crystallographic asymmetric unit, and the unit cell will be defined as having m crystallographic asymmetric units. Hence, there are $L = nm$ noncrystallographic asymmetric units within the unit cell. Clearly, the n noncrystallographic asymmetric units cannot completely fill the volume of one crystallographic asymmetric unit. The remaining space must be assumed to be empty or to be occupied by solvent molecules which disobey the noncrystallographic symmetry.

2.3.5.2. Interpretation of Pattersons in the presence of noncrystallographic symmetry

If noncrystallographic symmetry is present, an atom at a general position within the relevant volume will imply the presence of others within the same crystallographic asymmetric unit. If the noncrystallographic symmetry is known, then the positions of equivalent atoms may be generated from a single atomic position. The additional vector interactions which arise from crystallographically and noncrystallographically equivalent atoms in a crystal may be predicted and exploited in an interpretation of the Patterson function.

An object in real space which has a closed point group may incorporate some of its symmetry in the crystallographic symmetry. If there are l such objects in the cell, then there will be mn/l equivalent positions within each object. The 'self-vectors' formed between these positions within the object will be independent of the position of the objects. This distinction is important in that the self-vectors arising from atoms interacting with other atoms within a single particle may be correctly predicted without the knowledge of the particle centre position. In fact, this distinction may be exploited in a two-stage procedure in which an atom may be first located relative to the particle centre by use of the self-vectors and subsequently the particle may be positioned relative to crystallographic symmetry elements by use of the 'cross-vectors' (Table 2.3.5.1).

The interpretation of a heavy-atom difference Patterson for the holo-enzyme of lobster glyceraldehyde-3-phosphate dehydrogenase (GAPDH) provides an illustration of how the known noncrystallographic symmetry can aid the solution (Rossmann *et al.*, 1972; Buehner *et al.*, 1974). The GAPDH enzyme crystallized in a $P2_12_12_1$ cell (a = 149.0, b = 139.1, c = 80.7 Å) containing one tetramer per asymmetric unit. A rotation-function analysis had indicated the presence of three mutually perpendicular molecular twofold axes which suggested that the tetramer had 222 symmetry, and a locked rotation function determined the precise orientation of the tetramer relative to the crystal axes (see Table 2.3.5.2). Packing considerations led to assignment of a tentative particle centre near $\frac{1}{2}, \frac{1}{4}, Z$.

Table 2.3.5.1. *Possible types of vector searches*

Self-vectors	Cross-vectors	Dimension of search, n
(1) Locate single site relative to particle centre		$n = 3$
(2)	Use information from (1) to locate particle centre	$n \leq 3$
(3) Simultaneous search for both (1) and (2). In general this is a six-dimensional search but may be simplified when particle is on a crystallographic symmetry axis		$3 \leq n \leq 6$
(4) Given (1) for more than one site, find all vectors within particle		$n = 3$
(5) Given information from (3), locate additional site using complete vector set		$n = 3$

An isomorphous difference Patterson was calculated for the K_2HgI_4 derivative of GAPDH using data to a resolution of 6.8 Å. From an analysis of the three Harker sections, a tentative first heavy-atom position was assigned (atom A_2 at x, y, z). At this juncture, the known noncrystallographic symmetry was used to obtain a full interpretation. From Table 2.3.5.2 we see that molecular axis 2 will generate a second heavy atom with coordinates roughly $\frac{1}{4} + y, -\frac{1}{4} + x, 2Z - z$ (if the molecular centre was assumed to be at $\frac{1}{2}, \frac{1}{4}, Z$). Starting from the tentative coordinates of site A_2, the site A_1 related by molecular axis 1 was detected at about the predicted position and the second site A_1 generated acceptable cross-vectors with the earlier determined site A_2. Further examination enabled the completion of the set of four noncrystallographically related heavy-atom sites, such that all predicted Patterson vectors were acceptable and all four sites placed the molecular centre in the same position. Following refinement of these four sites, the corresponding SIR phases were used to find an additional set of four sites in this compound as well as in a number of other derivatives. The multiple isomorphous replacement phases, in conjunction with real-space electron-density averaging of the noncrystallographically related units, were then sufficient to solve the GAPDH structure.

When investigators studied larger macromolecular aggregates such as the icosahedral viruses, which have 532 point symmetry, systematic methods were developed for utilizing the noncrystallographic symmetry to aid in locating heavy-atom sites in isomorphous heavy-atom derivatives. Argos & Rossmann (1974, 1976) introduced an exhaustive Patterson search procedure for a single heavy-atom site within the noncrystallographic asymmetric unit which has been successfully applied to the interpretation of both virus [satellite tobacco necrosis virus (STNV) (Lentz *et al.*, 1976), southern bean mosaic virus (Rayment *et al.*, 1978), alfalfa mosaic virus (Fukuyama *et al.*, 1983), cowpea mosaic virus

Table 2.3.5.2. *Orientation of the glyceraldehyde-3-phosphate dehydrogenase molecular twofold axis in the orthorhombic cell*

Rotation axes	Polar coordinates (°)		Cartesian coordinates (direction cosines)		
	ψ	φ	u	v	w
1	45.0	−7.0	0.7018	0.7071	−0.0862
2	180.0–55.0	38.6	0.6402	−0.5736	0.5111
3	180.0–66.0	−70.6	0.3035	−0.4067	−0.8616

Table 2.3.6.1. *Different types of uses for the rotation function*

Type of rotation function	Pattersons to be compared		Purpose
	P_1	P_2	
Self	Unknown structure	Unknown structure, same cell	Finds orientation of noncrystallographic axes
Cross	Unknown structure	Unknown structure in different cell	Finds relative orientation of unknown molecules
Cross	Unknown structure	Known structure in large cell to avoid overlap of self-Patterson vectors	Determines orientation of unknown structure as preliminary to positioning and subsequent phasing with known molecule

(Stauffacher *et al.*, 1987)] and enzyme [catalase (Murthy *et al.*, 1981)] heavy-atom difference Pattersons. This procedure has also been implemented in the program *Patsol* (Tong & Rossmann, 1993). A heavy atom is placed in turn at all plausible positions within the volume of the noncrystallographic asymmetric unit and the corresponding vector set is constructed from the resulting constellation of heavy atoms. Argos & Rossmann (1976) found a spherical polar coordinate search grid to be convenient for spherical viruses. After all vectors for the current search position are predicted, the vectors are allocated to the nearest grid point and the list is sorted to eliminate recurring ones. The criterion used by Argos & Rossmann for selecting a solution is that the sum

$$S = \sum_{i=1}^{N} P_i - NP_{\rm av}$$

of the lookup Patterson density values P_i achieves a high value for a correct heavy-atom position. The sum is corrected for the carpet of cross-vectors by the second term in the sum.

An additional criterion, which has been found useful for discriminating correct solutions, is a unit vector density criterion (Arnold *et al.*, 1987)

$$U = \sum_{i=1}^{N} (P_i/n_i) \Big/ N,$$

where n_i is the number of vectors expected to contribute to the Patterson density value P_i (Arnold *et al.*, 1987). This criterion can be especially valuable for detecting correct solutions at special search positions, such as an icosahedral fivefold axis, where the number of vector lookup positions may be drastically reduced owing to the higher symmetry. An alternative, but equivalent, method for locating heavy-atom positions from isomorphous difference data is discussed in Section 2.3.3.5.

Even for a single heavy-atom site at a general position in the simplest icosahedral or ($T = 1$) virus, there are 60 equivalent heavy atoms in one virus particle. The number of unique vectors corresponding to this self-particle vector set will depend on the crystal symmetry but may be as many as $(60)(59)/2 = 1770$ for a virus particle at a general crystallographic position. Such was the case for the STNV crystals which were in space group $C2$ containing four virus particles at general positions. The method of Argos & Rossmann was applied successfully to a solution of the K_2HgI_4 derivative of STNV using a 10 Å resolution difference Patterson. Application of the noncrystallographic symmetry vector search procedure to a $K_2Au(CN)_2$ derivative of human rhinovirus 14 (HRV14) crystals (space group $P2_13$, $Z = 4$) has succeeded in establishing both the relative positions of heavy atoms within one particle and the positions of the virus particles relative to the crystal symmetry elements (Arnold *et al.*, 1987). The particle position was established by incorporating inter-particle vectors in the search and varying the particle position along the crystallographic threefold axis until the best fit for the predicted vector set was achieved.

Conversely, the knowledge that the heavy-atom positions, especially the Se atoms in a selenomethionyl protein, should obey the noncrystallographic symmetry can be used to deduce the nature, orientation and position of the noncrystallographic symmetry in the crystal unit cell, with either manual or automated procedures (Buehner *et al.*, 1974; Lu, 1999; Terwilliger, 2002*a*). The noncrystallographic symmetry can also serve as a powerful tool for refining the phase information derived from the heavy-atom positions (Buehner *et al.*, 1974).

2.3.6. Rotation functions

2.3.6.1. Introduction

The rotation function is designed to detect noncrystallographic rotational symmetry (see Table 2.3.6.1). The normal rotation function definition is given as (Rossmann & Blow, 1962)

$$R = \int_U P_1(\mathbf{u}) \cdot P_2(\mathbf{u}')\, {\rm d}\mathbf{u}, \qquad (2.3.6.1)$$

where P_1 and P_2 are two Pattersons and U is an envelope centred at the superimposed origins. This convolution therefore measures the degree of similarity, or 'overlap', between the two Pattersons when P_2 has been rotated relative to P_1 by an amount defined by

$$\mathbf{u}' = [\boldsymbol{C}]\mathbf{u}. \qquad (2.3.6.2)$$

The elements of $[\boldsymbol{C}]$ will depend on three rotation angles $(\theta_1, \theta_2, \theta_3)$. Thus, R is a function of these three angles. Alternatively, the matrix $[\boldsymbol{C}]$ could be used to express mirror symmetry, permitting searches for noncrystallographic mirror or glide planes.

The basic concepts were first clearly stated by Rossmann & Blow (1962), although intuitive uses of the rotation function had been considered earlier. Hoppe (1957*b*) had also hinted at a convolution of the type given by (2.3.6.1) to find the orientation of known molecular fragments and these ideas were implemented by Huber (1965).

Consider a structure of two identical units which are in different orientations. The Patterson function of such a structure consists of three parts. There will be the self-Patterson vectors of one unit, being the set of interatomic vectors which can be formed within that unit, with appropriate weights. The set of self-Patterson vectors of the other unit will be identical, but they will be rotated away from the first due to the different orientation. Finally, there will be the cross-Patterson vectors, or set of interatomic vectors which can be formed from one unit to another. The self-Patterson vectors of the two units will all lie in a volume centred at the origin and limited by the overall dimensions of the units. Some or all of the cross-Patterson vectors will lie outside this volume. Suppose the Patterson function is now superposed on a rotated version of itself. There will be no particular agreement except when one set of self-Patterson vectors of one unit has the same orientation as the self-Patterson vectors from the other unit. In this position, we would expect a maximum of agreement or 'overlap' between the two. Similarly,

the superposition of the molecular self-Patterson derived from different crystal forms can provide the relative orientation of the two crystals when the molecules are aligned.

While it would be possible to evaluate R by interpolating in P_2 and forming the point-by-point product with P_1 within the volume U for every combination of θ_1, θ_2 and θ_3, such a process is tedious and requires large computer storage for the Pattersons. Instead, the process is usually performed in reciprocal space where the number of independent structure amplitudes which form the Pattersons is about one-thirtieth of the number of Patterson grid points. Thus, the computation of a rotation function is carried out directly on the structure amplitudes, while the overlap definition (2.3.6.1) simply serves as a physical basis for the technique.

The derivation of the reciprocal-space expression depends on the expansion of each Patterson either as a Fourier summation, the conventional approach of Rossmann & Blow (1962), or as a sum of spherical harmonics in Crowther's (1972) analysis. The conventional and mathematically easier treatment is discussed presently, but the reader is referred also to Section 2.3.6.5 for Crowther's elegant approach. The latter leads to a rapid technique for performing the computations, about one hundred times faster than conventional methods.

Let, omitting constant coefficients,

$$P_1(\mathbf{u}) = \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{u})$$

and

$$P_2(\mathbf{u}') = \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 \exp(2\pi i \mathbf{p} \cdot \mathbf{u}').$$

From (2.3.6.2) it follows that

$$P_2(\mathbf{u}') = \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 \exp(2\pi i \mathbf{p}[\boldsymbol{C}] \cdot \mathbf{u}),$$

and, hence, by substitution in (2.3.6.1)

$$\begin{aligned} R(\theta_1, \theta_2, \theta_3) = \int_U & \left[\sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{u}) \right] \\ & \times \left[\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 \exp(2\pi i \mathbf{p}[\boldsymbol{C}] \cdot \mathbf{u}) \right] \mathrm{d}\mathbf{u} \\ = U \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}|^2 & \left(\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 G_{\mathbf{hp}} \right), \end{aligned} \tag{2.3.6.3}$$

where

$$U G_{\mathbf{hp}} = \int_U \exp\{2\pi i (\mathbf{h} + \mathbf{p}[\boldsymbol{C}]) \cdot \mathbf{u}\} \, \mathrm{d}\mathbf{u}.$$

When the volume U is a sphere, $G_{\mathbf{hp}}$ has the analytical form

$$G_{\mathbf{hp}} = \frac{3(\sin\theta - \theta\cos\theta)}{\theta^3}, \tag{2.3.6.4}$$

where $\theta = 2\pi HR$ and $H = \mathbf{h} + \mathbf{p}[\boldsymbol{C}]$. G is a spherical interference function whose form is shown in Fig. 2.3.6.1.

The expression (2.3.6.3) represents the rotation function in reciprocal space. If $\mathbf{h}' = [\boldsymbol{C}^T]\mathbf{p}$ in the argument of $G_{\mathbf{hp}}$, then $\mathbf{h}'$ can be seen as the point in reciprocal space to which $\mathbf{p}$ is rotated by

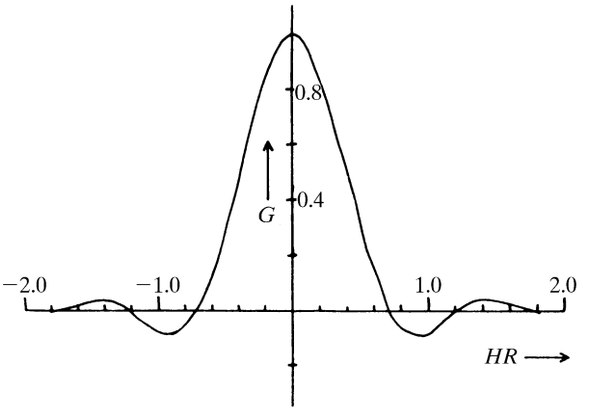


Fig. 2.3.6.1. Shape of the interference function G for a spherical envelope of radius R at a distance H from the reciprocal-space origin. [Reprinted from Rossmann & Blow (1962).]

$[\boldsymbol{C}]$. Only for those integral reciprocal-lattice points which are close to $-\mathbf{h}'$ will $G_{\mathbf{hp}}$ be of an appreciable size (Fig. 2.3.6.1). Thus, the number of significant terms is greatly reduced in the summation over $\mathbf{p}$ for every value of $\mathbf{h}$, making the computation of the rotation function manageable.

The radius of integration R should be approximately equal to or a little smaller than the molecular diameter. If R were roughly equal to the length of a lattice translation, then the separation of reciprocal-lattice points would be about $1/R$. Hence, when H is equal to one reciprocal-lattice separation, $HR \simeq 1$, and G is thus quite small. Indeed, all terms with $HR > 1$ might well be neglected. Thus, in general, the only terms that need be considered are those where $-\mathbf{h}'$ is within one lattice point of $\mathbf{h}$. However, in dealing with a small molecular fragment for which R is small compared to the unit-cell dimensions, more reciprocal-lattice points must be included for the summation over $\mathbf{p}$ in the rotation-function expression (2.3.6.3).

In practice, the equation

$$\mathbf{h} + \mathbf{h}' = 0,$$

that is

$$[\boldsymbol{C}^T]\mathbf{p} = -\mathbf{h}$$

or

$$\mathbf{p} = [\boldsymbol{C}^T]^{-1}(-\mathbf{h}), \tag{2.3.6.5}$$

determines $\mathbf{p}$, given a set of Miller indices $\mathbf{h}$. This will give a non-integral set of Miller indices. The terms included in the inner summation of (2.3.6.3) will be integral values of $\mathbf{p}$ around the non-integral lattice point found by solving (2.3.6.5).

Details of the conventional program were given by Tollin & Rossmann (1966) and follow the principles outlined above. They discussed various strategies as to which crystal should be used to calculate the first ($\mathbf{h}$) and second ($\mathbf{p}$) Patterson. Rossmann & Blow (1962) noted that the factor $\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}|^2 G_{\mathbf{hp}}$ in expression (2.3.6.3) represents an interpolation of the squared transform of the self-Patterson of the second ($\mathbf{p}$) crystal. Thus, the rotation function is a sum of the products of the two molecular transforms taken over all the $\mathbf{h}$ reciprocal-lattice points. Lattman & Love (1970) therefore computed the molecular transform explicitly and stored it in the computer, sampling it as required by the rotation operation. A discussion on the suitable choice of variables in the computation of rotation functions has been given by Lifchitz (1983).

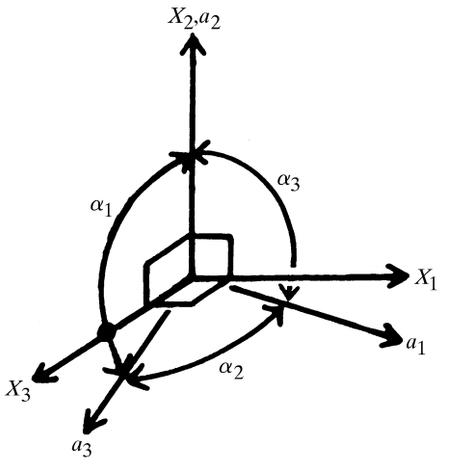


Fig. 2.3.6.2. Relationships of the orthogonal axes X_1, X_2, X_3 to the crystallographic axes a_1, a_2, a_3. [Reprinted from Rossmann & Blow (1962).]

2.3.6.2. Matrix algebra

The initial step in the rotation-function procedure involves the orthogonalization of both crystal systems. Thus, if fractional coordinates in the first crystal system are represented by **x**, these can be orthogonalized by a matrix $[\boldsymbol{\beta}]$ to give the coordinates **X** in units of length (Fig. 2.3.6.2); that is,

$$\mathbf{X} = [\boldsymbol{\beta}]\mathbf{x}.$$

If the point **X** is rotated to the point $\mathbf{X}'$, then

$$\mathbf{X}' = [\boldsymbol{\rho}]\mathbf{X}, \qquad (2.3.6.6)$$

where $\boldsymbol{\rho}$ represents the rotation matrix relating the two vectors in the orthogonal system. Finally, $\mathbf{X}'$ is converted back to fractional coordinates measured along the oblique cell dimension in the second crystal by

$$\mathbf{x}' = [\boldsymbol{\alpha}]\mathbf{X}'.$$

Thus, by substitution,

$$\mathbf{x}' = [\boldsymbol{\alpha}][\boldsymbol{\rho}]\mathbf{X} = [\boldsymbol{\alpha}][\boldsymbol{\rho}][\boldsymbol{\beta}]\mathbf{x}, \qquad (2.3.6.7)$$

and by comparison with (2.3.6.2) it follows that

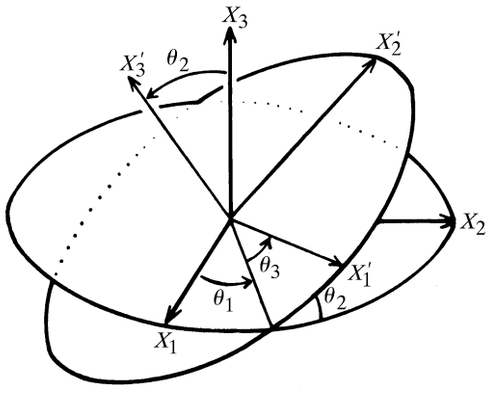


Fig. 2.3.6.3. Eulerian angles $\theta_1, \theta_2, \theta_3$ relating the rotated axes X_1', X_2', X_3' to the original unrotated orthogonal axes X_1, X_2, X_3. [Reprinted from Rossmann & Blow (1962).]

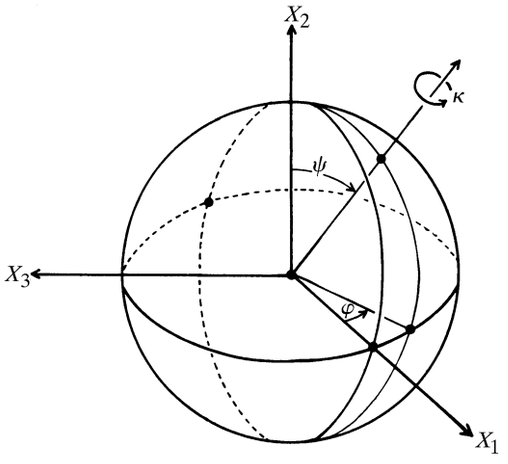


Fig. 2.3.6.4. Variables ψ and φ are polar coordinates which specify a direction about which the axes may be rotated through an angle κ. [Reprinted from Rossmann & Blow (1962).]

$$[\boldsymbol{C}] = [\boldsymbol{\alpha}][\boldsymbol{\rho}][\boldsymbol{\beta}].$$

Fig. 2.3.6.2 shows the mode of orthogonalization used by Rossmann & Blow (1962). With their definition it can be shown that

$$[\boldsymbol{\alpha}] = \begin{pmatrix} 1/(a_1 \sin\alpha_3 \sin\omega) & 0 & 0 \\ 1/(a_2 \tan\alpha_1 \tan\omega) & 1/a_2 & -1/(a_2 \tan\alpha_1) \\ -1/(a_2 \tan\alpha_3 \sin\omega) & & \\ -1/(a_3 \sin\alpha_1 \tan\omega) & 0 & 1/(a_3 \sin\alpha_1) \end{pmatrix}$$

and

$$[\boldsymbol{\beta}] = \begin{pmatrix} a_1 \sin\alpha_3 \sin\omega & 0 & 0 \\ a_1 \cos\alpha_3 & a_2 & a_3 \cos\alpha_1 \\ a_1 \sin\alpha_3 \cos\omega & 0 & a_3 \sin\alpha_1 \end{pmatrix},$$

where $\cos\omega = (\cos\alpha_2 - \cos\alpha_1 \cos\alpha_3)/(\sin\alpha_1 \sin\alpha_3)$ with $0 \le \omega < \pi$. For a Patterson compared with itself, $[\boldsymbol{\alpha}] = [\boldsymbol{\beta}]^{-1}$.

An alternative mode of orthogonalization, used by the Protein Data Bank and most programs, is to align the a_1 axis of the unit cell with the Cartesian X_1 axis, and to align the a_3^* axis with the Cartesian X_3 axis. With this definition, the orthogonalization matrix is

$$[\boldsymbol{\beta}] = \begin{pmatrix} a_1 & a_2 \cos\alpha_3 & a_3 \cos\alpha_2 \\ 0 & a_2 \sin\alpha_3 & -a_3 \sin\alpha_2 \cos\alpha_1^* \\ 0 & 0 & a_3 \sin\alpha_2 \sin\alpha_1^* \end{pmatrix}.$$

Other modes of orthogonalization are also possible, some of which are supported in the program *GLRF* (Tong & Rossmann, 1990, 1997).

Both spherical (κ, ψ, φ) and Eulerian $(\theta_1, \theta_2, \theta_3)$ angles are used in evaluating the rotation function. The usual definitions employed are given diagrammatically in Figs. 2.3.6.3 and 2.3.6.4. They give rise to the following rotation matrices.

(*a*) Matrix $[\boldsymbol{\rho}]$ in terms of Eulerian angles $\theta_1, \theta_2, \theta_3$:

$$\begin{pmatrix} -\sin\theta_1\cos\theta_2\sin\theta_3 + \cos\theta_1\cos\theta_3 & \cos\theta_1\cos\theta_2\sin\theta_3 + \sin\theta_1\cos\theta_3 & \sin\theta_2\sin\theta_3 \\ -\sin\theta_1\cos\theta_2\cos\theta_3 - \cos\theta_1\sin\theta_3 & \cos\theta_1\cos\theta_2\cos\theta_3 - \sin\theta_1\sin\theta_3 & \sin\theta_2\cos\theta_3 \\ \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

and (b) matrix $[\boldsymbol{\rho}]$ in terms of rotation angle κ and the spherical polar coordinates ψ, φ:

$$\begin{pmatrix} \cos\kappa + \sin^2\psi\cos^2\varphi(1-\cos\kappa) & \sin\psi\cos\psi\cos\varphi(1-\cos\kappa) + \sin\psi\sin\varphi\sin\kappa & -\sin^2\psi\cos\varphi\sin\varphi(1-\cos\kappa) + \cos\psi\sin\kappa \\ \sin\psi\cos\psi\cos\varphi(1-\cos\kappa) - \sin\psi\sin\varphi\sin\kappa & \cos\kappa + \cos^2\psi(1-\cos\kappa) & -\sin\psi\cos\psi\sin\varphi(1-\cos\kappa) - \sin\psi\cos\varphi\sin\kappa \\ -\sin^2\psi\sin\varphi\cos\varphi(1-\cos\kappa) - \cos\psi\sin\kappa & -\sin\psi\cos\psi\sin\varphi(1-\cos\kappa) + \sin\psi\cos\varphi\sin\kappa & \cos\kappa + \sin^2\psi\sin^2\varphi(1-\cos\kappa) \end{pmatrix}$$

Alternatively, (b) can be expressed as

$$\begin{pmatrix} \cos\kappa + u^2(1-\cos\kappa) & uv(1-\cos\kappa) - w\sin\kappa & uw(1-\cos\kappa) + v\sin\kappa \\ vu(1-\cos\kappa) + w\sin\kappa & \cos\kappa + v^2(1-\cos\kappa) & uw(1-\cos\kappa) - u\sin\kappa \\ wu(1-\cos\kappa) - v\sin\kappa & wv(1-\cos\kappa) + u\sin\kappa & \cos\kappa + w^2(1-\cos\kappa) \end{pmatrix},$$

where u, v and w are the direction cosines of the rotation axis given by

$$\begin{aligned} u &= \sin\psi\cos\varphi, \\ v &= \cos\psi, \\ w &= -\sin\psi\sin\varphi. \end{aligned}$$

This latter form also demonstrates that the trace of a rotation matrix is $2\cos\kappa + 1$.

The relationship between the two sets of variables established by comparison of the elements of the two matrices yields

$$\begin{aligned} \cos(\kappa/2) &= \cos(\theta_2/2)\cos\left(\frac{\theta_1+\theta_3}{2}\right), \\ \tan\varphi &= -\cot(\theta_2/2)\sin\left(\frac{\theta_1+\theta_3}{2}\right)\sec\left(\frac{\theta_1-\theta_3}{2}\right), \\ \cos\varphi\tan\psi &= \cot\left(\frac{\theta_1-\theta_3}{2}\right). \end{aligned}$$

Since φ and ψ can always be chosen in the range 0 to π, these equations suffice to find (κ, ψ, φ) from any set $(\theta_1, \theta_2, \theta_3)$.

Another definition for the polar angles is also commonly used. In this definition, the angle ψ is measured from the Cartesian Z (X_3) axis, instead of the Y (X_2) axis. As most space groups have the unique axis along a_3, the ψ angle will define the inclination relative to the unique axis of the space group with this definition.

2.3.6.3. Symmetry

In analogy with crystal lattices, the rotation function is periodic and contains symmetry. The rotation function has a cell whose periodicity is 2π in each of its three angles. This may be written as

$$R(\theta_1, \theta_2, \theta_3) \equiv R(\theta_1 + 2\pi n_1, \theta_2 + 2\pi n_2, \theta_3 + 2\pi n_3)$$

or

$$R(\kappa, \psi, \varphi) \equiv R(\kappa + 2\pi n_1, \psi + 2\pi n_2, \varphi + 2\pi n_3),$$

where n_1, n_2 and n_3 are integers. A redundancy in the definition of either set of angles leads to the equivalence of the following points:

$$R(\theta_1, \theta_2, \theta_3) \equiv R(\theta_1 + \pi, -\theta_2, \theta_3 + \pi) \quad \text{in Eulerian space}$$

or

$$R(\kappa, \psi, \varphi) \equiv R(\kappa, 2\pi - \psi, \varphi + \pi) \quad \text{in polar space.}$$

These relationships imply an n glide plane perpendicular to θ_2 for Eulerian space or a φ glide plane perpendicular to ψ in polar space.

In addition, the Laue symmetry of the two Pattersons themselves must be considered. This problem was first discussed by Rossmann & Blow (1962) and later systematized by Tollin *et al.* (1966), Burdina (1970, 1971, 1973) and Rao *et al.* (1980). A closely related problem was considered by Hirshfeld (1968). The rotation function will have the same value whether the Patterson density at $\mathbf{X}$ or $[\boldsymbol{T}_i]\mathbf{X}$ in the first crystal is multiplied by the Patterson density at $\mathbf{X}'$ or $[\boldsymbol{T}_j]\mathbf{X}'$ in the second crystal. $[\boldsymbol{T}_i]$ and $[\boldsymbol{T}_j]$ refer to the ith and jth crystallographic rotations in the orthogonalized coordinate systems of the first and second crystal, respectively. Hence, from (2.3.6.6)

$$([\boldsymbol{T}_j]\mathbf{X}') = [\boldsymbol{\rho}]([\boldsymbol{T}_i]\mathbf{X})$$

or

$$\mathbf{X}' = [\boldsymbol{T}_j^T][\boldsymbol{\rho}][\boldsymbol{T}_i]\mathbf{X}.$$

Thus, it is necessary to find angular relationships which satisfy the relation

$$[\boldsymbol{\rho}] = [\boldsymbol{T}_j^T][\boldsymbol{\rho}][\boldsymbol{T}_i]$$

for given Patterson symmetries. Tollin *et al.* (1966) show that the Eulerian angular equivalences can be expressed in terms of the Laue symmetries of each Patterson (Table 2.3.6.2).

The example given by Tollin *et al.* (1966) is instructive in the use of Table 2.3.6.2. They consider the determination of the Eulerian space group when P_1 has symmetry *Pmmm* and P_2 has symmetry $P2/m$. These Pattersons contain the proper rotation groups 222 and 2 (parallel to **b**), respectively. Inspection of Table 2.3.6.2 shows that these symmetries produce the following Eulerian relationships:

Table 2.3.6.2. *Eulerian symmetry elements for all possible types of space-group rotations*

Axis	Direction	First crystal	Second crystal
1		$(\pi+\theta_1, -\theta_2, \pi+\theta_3)$	$(\pi+\theta_1, -\theta_2, \pi+\theta_3)$
2	[010]	$(\pi-\theta_1, \pi+\theta_2, \theta_3)$	$(\theta_1, \pi+\theta_2, \pi-\theta_3)$
2	[001]	$(\pi+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, \pi+\theta_3)$
4	[001]	$(-\pi/2+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, \pi/2+\theta_3)$
3	[001]	$(-2\pi/3+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, 2\pi/3+\theta_3)$
6	[001]	$(-\pi/3+\theta_1, \theta_2, \theta_3)$	$(\theta_1, \theta_2, \pi/3+\theta_3)$
2†	[110]	$(3\pi/2-\theta_1, \pi-\theta_2, \pi+\theta_3)$	$(\pi+\theta_1, \pi-\theta_2, -3\pi/2-\theta_3)$

† This axis is not unique (that is, it can always be generated by two other unique axes), but is included for completeness.

Table 2.3.6.3. *Numbering of the rotation-function space groups*

The Laue group of the rotated Patterson map P_1 is chosen from the left column and the Laue group of P_2 is chosen from the upper row.

	1	2/*m*, *b* axis unique	2/*m*, *c* axis unique	*mmm*	4/*m*	4/*mmm*	3	3*m*	6/*m*	6/*mmm*
1	1	11	21	31	41	51	61	71	81	91
2/*m*, *b* axis unique	2	12	22	32	42	52	62	72	82	92
2/*m*, *c* axis unique	3	13	23	33	43	53	63	73	83	93
mmm	4	14	24	34	44	54	64	74	84	94
4/*m*	5	15	25	35	45	55	65	75	85	95
4/*mmm*	6	16	26	36	46	56	66	76	86	96
3	7	17	27	37	47	57	67	77	87	97
3*m*	8	18	28	38	48	58	68	78	88	98
6/*m*	9	19	29	39	49	59	69	79	89	99
6/*mmm*	10	20	30	40	50	60	70	80	90	100

(*a*) In the first crystal (*Pmmm*):

$\theta_1\theta_2\theta_3 \rightarrow \pi+\theta_1, -\theta_2, \pi+\theta_3$ (onefold axis)
$\theta_1\theta_2\theta_3 \rightarrow \pi-\theta_1, \pi+\theta_2, \theta_3$ (twofold axis parallel to **b**)
$\theta_1\theta_2\theta_3 \rightarrow \pi+\theta_1, \theta_2, \theta_3$ (twofold axis parallel to **c**).

(*b*) In the second crystal (*P*2/*m*):

$\theta_1\theta_2\theta_3 \rightarrow \pi+\theta_1, -\theta_2, \pi+\theta_3$ (onefold axis)
$\theta_1\theta_2\theta_3 \rightarrow \theta_1, \pi+\theta_2, \pi-\theta_3$ (twofold axis parallel to **b**).

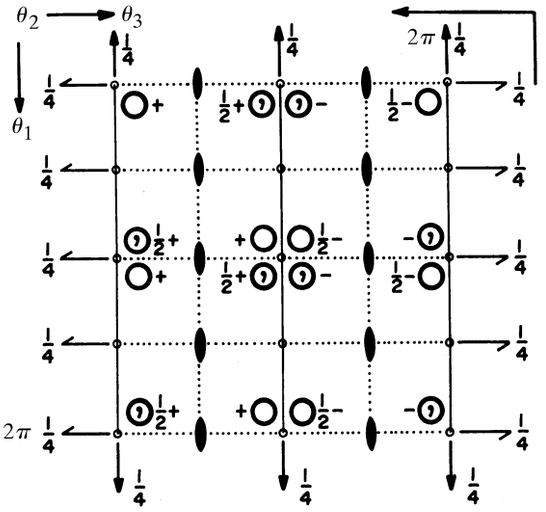


Fig. 2.3.6.5. Rotation space group diagram for the rotation function of a *Pmmm* Patterson function (P_1) against a *P*2/*m* Patterson function (P_2). The Eulerian angles $\theta_1, \theta_2, \theta_3$ repeat themselves after an interval of 2π. Heights above the plane are given in fractions of a revolution. [Reprinted from Tollin *et al.* (1966).]

When these symmetry operators are combined two cells result, each of which has the space group *Pbcb* (Fig. 2.3.6.5). The asymmetric unit within which the rotation function need be evaluated is found from a knowledge of the Eulerian space group. In the above example, the limits of the asymmetric unit are $0 \le \theta_1 \le \pi/2$, $0 \le \theta_2 \le \pi$ and $0 \le \theta_3 \le \pi/2$.

Nonlinear transformations occur when using Eulerian symmetries for threefold axes along [111] (as in the cubic system) or when using polar coordinates. Hence, Eulerian angles are far more suitable for a derivation of the limits of the rotation-function asymmetric unit. However, when searching for given molecular axes, where some plane of κ need be explored, polar angles are more useful.

Rao *et al.* (1980) have determined all possible rotation-function Eulerian space groups, except for combinations with Pattersons of cubic space groups. They numbered these rotation groups 1 through 100 (Table 2.3.6.3) according to the combination of the Patterson Laue groups. The characteristics of each of the 100 groups are given in Table 2.3.6.4, including the limits of the asymmetric unit. In the 100 unique combinations of noncubic Laue groups, there are only 16 basic rotation-function space groups.

2.3.6.4. Sampling, background and interpretation

If the origins are retained in the Pattersons, their product will form a high but constant plateau on which the rotation-function peaks are superimposed; this leads to a small apparent peak-to-noise ratio. The effect can be eliminated by removal of the origins through a modification of the Patterson coefficients. Irrespective of origin removal, a significant peak is one which is more than three r.m.s. deviations from the mean background.

As in all continuous functions sampled at discrete points, a convenient grid size must be chosen. Small intervals result in an excessive computing burden, while large intervals might miss peaks. Furthermore, equal increments of angles do not represent equal changes in rotation, which can result in distorted peaks (Lattman, 1972). In general, a crude idea of a useful sampling

Table 2.3.6.4. *Rotation-function Eulerian space groups*

The rotation space groups are given in Table 2.3.6.3.

No. of the rotation space group	No. of equivalent positions[(a)]	Symbol[(b)]	Translation along the θ_1 axis[(c)]	Translation along the θ_3 axis[(c)]	Range of the asymmetric unit[(d)]		
1	2	*Pn*	2π	2π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
2	4	$Pbn2_1$	2π	2π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
3	4	*Pc*	π	2π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
4	8	$Pbc2_1$	π	2π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
5	8	*Pc*	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
6	16	$Pbc2_1$	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
7	6	*Pn*	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
8	12	$Pbn2_1$	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
9	12	*Pc*	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi$
10	24	$Pbc2_1$	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
11	4	$P2_1nb$	2π	2π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
12	8	*Pbnb*	2π	2π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < 2\pi$
13	8	*P2cb*	π	2π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
14	16	*Pbcb*	π	2π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
15	16	*P2cb*	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
16	32	*Pbcb*	$\pi/2$	2π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 \le \pi/2$
17	12	$P2_1nb$	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
18	24	*Pbnb*	$2\pi/3$	2π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 \le \pi/2$
19	24	*P2cb*	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi$
20	48	*Pbcb*	$\pi/3$	2π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 \le \pi/2$
21	4	*Pa*	2π	π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
22	8	*Pba2*	2π	π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
23	8	*Pm*	π	π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
24	16	*Pbm2*	π	π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
25	16	*Pm*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
26	32	*Pbm2*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
27	12	*Pa*	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
28	24	*Pba2*	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
29	24	*Pm*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi$
30	48	*Pbm2*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
31	8	$P2_1ab$	2π	π	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
32	16	*Pbab*	2π	π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < \pi$
33	16	*P2mb*	π	π	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
34	32	*Pbmb*	π	π	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
35	32	*P2mb*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
36	64	*Pbmb*	$\pi/2$	π	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
37	24	$P2_1ab$	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
38	48	*Pbab*	$2\pi/3$	π	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
39	48	*P2mb*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi$
40	96	*Pbmb*	$\pi/3$	π	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
41	8	*Pa*	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
42	16	*Pba2*	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
43	16	*Pm*	π	$\pi/2$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
44	32	*Pbm2*	π	$\pi/2$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
45	32	*Pm*	$\pi/2$	$\pi/2$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
46	64	*Pbm2*	$\pi/2$	$\pi/2$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
47	24	*Pa*	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
48	48	*Pba2*	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
49	48	*Pm*	$\pi/3$	$\pi/2$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/2$
50	96	*Pbm2*	$\pi/3$	$\pi/2$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
51	16	$P2_1ab$	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
52	32	*Pbab*	2π	$\pi/2$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/4$
53	32	*P2mb*	π	$\pi/2$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
54	64	*Pbmb*	π	$\pi/2$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
55	64	*P2mb*	$\pi/2$	$\pi/2$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
56	128	*Pbmb*	$\pi/2$	$\pi/2$	$0 \le \theta_1 \le \pi/4$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
57	48	$P2_1ab$	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
58	96	*Pbab*	$2\pi/3$	$\pi/2$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/4$

Table 2.3.6.4 (*cont.*)

No. of the rotation space group	No. of equivalent positions[(a)]	Symbol[(b)]	Translation along the θ_1 axis[(c)]	Translation along the θ_3 axis[(c)]	Range of the asymmetric unit[(d)]		
59	96	*P2mb*	$\pi/3$	$\pi/2$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
60	192	*Pbmb*	$\pi/3$	$\pi/2$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/2$
61	6	*Pn*	2π	$2\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
62	12	$Pbn2_1$	2π	$2\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
63	12	*Pc*	π	$2\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
64	24	$Pbc2_1$	π	$2\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
65	24	*Pc*	$\pi/2$	$2\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
66	48	$Pbc2_1$	$\pi/2$	$2\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
67	18	*Pn*	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
68	36	$Pbn2_1$	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
69	36	*Pc*	$\pi/3$	$2\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
70	72	$Pbc2_1$	$\pi/3$	$2\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
71	12	$P2_1nb$	2π	$2\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
72	24	*Pbnb*	2π	$2\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < 2\pi/3$
73	24	*P2cb*	π	$2\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
74	48	*Pbcb*	π	$2\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
75	48	*P2cb*	$\pi/2$	$2\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
76	96	*Pbcb*	$\pi/2$	$2\pi/3$	$0 \le \theta_1 \le \pi/4$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
77	36	$P2_1nb$	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
78	72	*Pbnb*	$2\pi/3$	$2\pi/3$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < 2\pi/3$
79	72	*P2 cb*	$\pi/3$	$2\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
80	144	*Pbcb*	$\pi/3$	$2\pi/3$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < 2\pi/3$
81	12	*Pa*	2π	$\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
82	24	*Pba2*	2π	$\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
83	24	*Pm*	π	$\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
84	48	*Pbm2*	π	$\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
85	48	*Pm*	$\pi/2$	$\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
86	96	*Pbm2*	$\pi/2$	$\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
87	36	*Pa*	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
88	72	*Pba2*	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
89	72	*Pm*	$\pi/3$	$\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi$,	$0 \le \theta_3 < \pi/3$
90	144	*Pbm2*	$\pi/3$	$\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
91	24	$P2_1ab$	2π	$\pi/3$	$0 \le \theta_1 < 2\pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
92	48	*Pbab*	2π	$\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 < \pi$,	$0 \le \theta_3 < \pi/3$
93	48	*P2mb*	π	$\pi/3$	$0 \le \theta_1 < \pi$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
94	96	*Pbmb*	π	$\pi/3$	$0 \le \theta_1 \le \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/2$
95	96	*P2mb*	$\pi/2$	$\pi/3$	$0 \le \theta_1 < \pi/2$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
96	192	*Pbmb*	$\pi/2$	$\pi/3$	$0 \le \theta_1 \le \pi/4$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
97	72	$P2_1ab$	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
98	144	*Pbab*	$2\pi/3$	$\pi/3$	$0 \le \theta_1 < 2\pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 \le \pi/6$
99	144	*P2mb*	$\pi/3$	$\pi/3$	$0 \le \theta_1 < \pi/3$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$
100	288	*Pbmb*	$\pi/3$	$\pi/3$	$0 \le \theta_1 \le \pi/6$,	$0 \le \theta_2 \le \pi/2$,	$0 \le \theta_3 < \pi/3$

Notes: (*a*) This is the number of equivalent positions in the rotation unit cell. (*b*) Each symbol retains the order $\theta_1, \theta_2, \theta_3$. The monoclinic space groups have the *b* axis unique setting. (*c*) This is a translation symmetry: *e.g.* for the case of $\pi/2$ translation along the θ_1 axis, $\theta_1, \theta_2, \theta_3$ goes to $\pi/2 + \theta_1, \theta_2, \theta_3$ and $\pi + \theta_1, \theta_2, \theta_3$, and $3\pi/2 + \theta_1, \theta_2, \theta_3$. All other equivalent positions in the basic rotation space group are similarly translated. (*d*) Several consistent sets of ranges exist but the one with the minimum range of θ_2 is listed.

interval can be obtained by considering the angle necessary to move one reciprocal-lattice point onto its neighbour (separated by a^*) at the extremity of the resolution limit, R. This interval is given by

$$\Delta\theta = a^*/2(1/R) = \tfrac{1}{2}Ra^*.$$

Simple sharpening of the rotation function can be useful. This can be achieved by restricting the computations to a shell in reciprocal space or by using normalized structure factors. Useful limits are frequently 10 to 6, 10 to 4 or 10 to 3.5 Å for an average protein or 6 to 5 Å for a virus structure determination. In addition, use of restricted resolution ranges, such as 6 to 5 Å or 3.5 to 3.0 Å, has been found in numerous cases to give especially well defined results (Arnold *et al.*, 1984).

When exploring the rotation function in polar coordinates, there is no significance to the latitude φ (Fig. 2.3.6.4) when $\psi = 0$. For small values of ψ, the rotation function will be quite insensitive to φ, which therefore needs to be explored only at coarse intervals (say 45°). As ψ approaches the equator at 90°, optimal increments of ψ and φ will be about equal. A similar situation exists with Eulerian angles. When $\theta_2 = 0$, the rotation function will be determined by $\theta_1 + \theta_3$, corresponding to $\psi = 0$ and

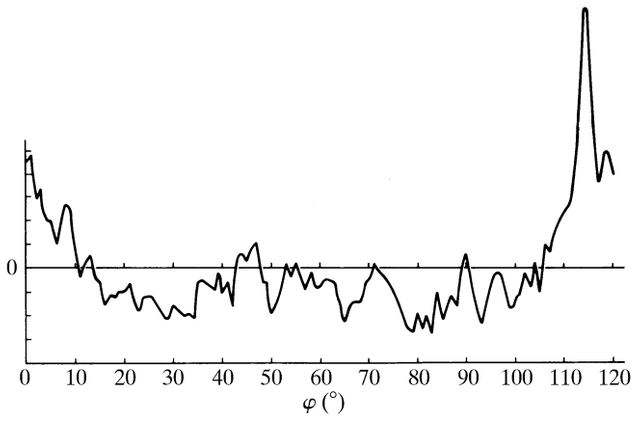


Fig. 2.3.6.6. The locked rotation function, L, applied to the determination of the orientation of the common cold virus (Arnold *et al.*, 1984). There are four virus particles per cubic cell with each particle sitting on a threefold axis. The locked rotation function explores all positions of rotation about this axis and, hence, repeats itself after 120°. The locked rotation function is determined from the individual rotation-function values of the noncrystallographic symmetry directions of a 532 icosahedron. [Reprinted with permission from Arnold *et al.* (1984).]

varying κ in polar coordinates. There will be no dependence on $(\theta_1 - \theta_3)$. Thus Eulerian searches can often be performed more economically in terms of the variables $\eta = \theta_1 + \theta_3$ and $\Delta = \theta_1 - \theta_3$, where

$$[\boldsymbol{\rho}] = \begin{pmatrix} \left[\cos\eta\cos^2\left(\frac{\theta_2}{2}\right) + \cos\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \left[\sin\eta\cos^2\left(\frac{\theta_2}{2}\right) + \sin\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \sin\theta_2\sin(\eta-\Delta) \\ \left[-\sin\eta\cos^2\left(\frac{\theta_2}{2}\right) + \sin\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \left[\cos\eta\cos^2\left(\frac{\theta_2}{2}\right) - \cos\Delta\sin^2\left(\frac{\theta_2}{2}\right)\right] & \sin\theta_2\cos(\eta-\Delta) \\ \sin\theta_2\sin(\eta+\Delta) & -\sin\theta_2\cos(\eta+\Delta) & \cos\theta_2 \end{pmatrix},$$

which reduces to the simple rotation matrix

$$[\boldsymbol{\rho}] = \begin{pmatrix} \cos\eta & \sin\eta & 0 \\ -\sin\eta & \cos\eta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

when $\theta_2 = 0$.

The computational effort to explore carefully a complete asymmetric unit of the rotation-function Eulerian group can be considerable. However, unless improper rotations are under investigation (as, for example, cross-rotation functions between different crystal forms of the same molecule), it is not generally necessary to perform such a global search. The number of molecules per crystallographic asymmetric unit, or the number of subunits per molecule, are often good indicators as to the possible types of noncrystallographic symmetry element. For instance, in the early investigation of insulin, the rotation function was used to explore only the $\kappa = 180°$ plane in polar coordinates as there were only two molecules per crystallographic asymmetric unit (Dodson *et al.*, 1966). Rotation functions of viruses, containing 532 icosahedral symmetry, are usually limited to exploration of the κ = 180, 120, 72 and 144° planes [*e.g.* Rayment *et al.* (1978) and Arnold *et al.* (1984)].

In general, the interpretation of the rotation function is straightforward. However, noise often builds up relative to the signal in high-symmetry space groups or if the data are limited or poor. One aid to a systematic interpretation is the locked rotation function (Rossmann *et al.*, 1972) for use when a molecule has more than one noncrystallographic symmetry axis. It is then possible to determine the rotation-function values for each molecular axis for a chosen molecular orientation (Fig. 2.3.6.6) (see Section 2.3.6.6).

Another problem in the interpretation of rotation functions is the appearance of apparent noncrystallographic symmetry that relates the self-Patterson of one molecule to the self-Patterson of a crystallographically related molecule. For example, take the case of α-chymotrypsin (Blow *et al.*, 1964). The space group is $P2_1$ with a molecular dimer in each of the two crystallographic asymmetric units. The noncrystallographic dimer axis was found to be perpendicular to the crystallographic 2_1 axis. The product of the crystallographic twofold in the Patterson with the orthogonal twofold in the self-Patterson vectors around the origin creates a third twofold, orthogonal to both of the other twofolds. In real space this represents a twofold screw direction relating the two dimers in the cell. In other cases, the product of the crystallographic and noncrystallographic symmetry results in symmetry which only has meaning in terms of all the vectors in the vicinity of the Patterson origin, but not in real space. Rotation-function peaks arising from such products are called Klug peaks (Johnson *et al.*, 1975). Such peaks normally refer to the total symmetry of all the vectors around the Patterson origin and may, therefore, be much larger than the peaks due to noncrystallographic symmetry within one molecule alone. Hence the Klug peaks, if not correctly recognized, can lead to erroneous conclusions (Åkervall *et al.*, 1972). Litvin (1975) has shown how Klug peaks can be predicted. These usually occur only for special orientations of a particle with a given symmetry relative to the crystallographic symmetry axes. Prediction of Klug peaks requires the simultaneous consideration of the noncrystallographic point group, the crystallographic point group and their relative orientations.

A special, but frequently occurring, situation arises when an evenfold noncrystallographic symmetry operator (*e.g.* 2-, 4-, 6-, 8-*etc.* fold axes) is parallel, or nearly parallel, to a crystallographic evenfold axis or screw axis. If the crystallographic evenfold axis is, say, parallel to Z, then if the centre of molecule I is at (X_0, Y_0, Z_0), the centre of molecule II will be at $(-X_0, -Y_0, Z_0)$. If molecule I has an evenfold axis parallel to Z, then for every atom (a) at $(x + X_0, y + Y_0, z + Z_0)$, there will be an atom (b) at $(-x + X_0, -y + Y_0, z + Z_0)$. The crystallographic symmetry-equivalent positions of these two atoms in molecule II will be at (c) $(-x - X_0, -y - Y_0, z + Z_0)$ and (d) $(x - X_0, y - Y_0, z + Z_0)$. The vectors between atoms (a) and (d) and also between atoms (b) and (c) will both have component lengths of $(2X_0, 2Y_0, 0)$. The position of this vector in a Patterson map is independent of the actual atoms in the molecule and depends only on the position of the molecular noncrystallographic symmetry axis.

Every atom will produce two vectors of this type, all of which will accumulate in a Patterson map to produce a large peak, which establishes the exact position of the noncrystallographic symmetry evenfold axis relative to the crystallographic axis. The position of the special peak is on the Harker section, namely at w = 0 for a crystallographic twofold axis and at $w = \frac{1}{2}$ for a crystallographic 2_1 screw axis. If there are N atoms in the structure of the two crystallographically related dimers, then the height of the origin is proportional to N (the number of zero-length vectors). The number of vectors with length $(2X_0, 2Y_0, 0)$ will be twice the number of atoms in each monomer, or $2 \times (N/4)$, which is $N/2$. Thus the special peak should be about half the height of the Patterson map's origin peak. In practice, the peak is often somewhat lower because the noncrystallographic symmetry and crystallographic axes might not be exactly parallel. This situation can be mitigated by computing the Patterson map with lower-resolution reflections only, as the difference in orientation between the axes is less significant when viewed at lower resolution (McKenna *et al.*, 1992).

2.3.6.5. The fast rotation function

Unfortunately, the rotation-function computations can be extremely time-consuming by conventional methods. Sasada (1964) developed a technique for rapidly finding the maximum of a given peak by looking at the slope of the rotation function. A major breakthrough came when Crowther (1972) recast the rotation function in a manner suitable for rapid computation. Only a brief outline of Crowther's fast rotation function is given here. Details are found in the original text (Crowther, 1972) and his computer program description.

Since the rotation function correlates spherical volumes of a given Patterson density with rotated versions of either itself or another Patterson density, it is likely that a more natural form for the rotation function will involve spherical harmonics rather than the Fourier components $|\mathbf{F_h}|^2$ of the crystal representation. Thus, if the two Patterson densities $P_1(r, \psi, \varphi)$ and $P_2(r, \psi, \varphi)$ are expanded within the spherical volume of radius less than a limiting value of a, then

$$P_1(r, \psi, \varphi) = \sum_{lmn} a^*_{lmn} \hat{j}_l(k_{ln} r) \hat{Y}_l^{m^*}(\psi, \varphi)$$

and

$$P_2(r, \psi, \varphi) = \sum_{l'm'n'} b_{l'm'n'} \hat{j}_{l'}(k_{l'n'} r) \hat{Y}_{l'}^{m'}(\psi, \varphi),$$

and the rotation function would then be defined as

$$R = \int_{\text{sphere}} P_1(r, \psi, \varphi) \mathscr{R} P_2(r, \psi, \varphi) r^2 \sin\psi \, \mathrm{d}r \, \mathrm{d}\psi \, \mathrm{d}\varphi.$$

Here $\hat{Y}_l^m(\psi, \varphi)$ is the normalized spherical harmonic of order l; $\hat{j}_l(k_{ln} r)$ is the normalized spherical Bessel function of order l; a_{lmn}, b_{lmn} are complex coefficients; and $\mathscr{R}P_2(r, \psi, \varphi)$ represents the rotated second Patterson. The rotated spherical harmonic can then be expressed in terms of the Eulerian angles $\theta_1, \theta_2, \theta_3$ as

$$\mathscr{R}(\theta_1, \theta_2, \theta_3) \hat{Y}_l^m(\psi, \varphi) = \sum_{q=-l}^{l} D^l_{qm}(\theta_1, \theta_2, \theta_3) \hat{Y}_l^q(\psi, \varphi),$$

where

$$D^l_{qm}(\theta_1, \theta_2, \theta_3) = \exp(iq\theta_3) d^l_{qm}(\theta_2) \exp(im\theta_1)$$

and $d^l_{qm}(\theta_2)$ are the matrix elements of the three-dimensional rotation group. It can then be shown that

$$R(\theta_1, \theta_2, \theta_3) = \sum_{lmm'n} a^*_{lmn} b_{lm'n} D^l_{m'm}(\theta_1, \theta_2, \theta_3).$$

Since the radial summation over n is independent of the rotation,

$$c_{lmm'} = \sum_n a^*_{lmn} b_{lmn},$$

and hence

$$R(\theta_1, \theta_2, \theta_3) = \sum_{lmm'} c_{lmm'} D^l_{m'm}(\theta_1, \theta_2, \theta_3)$$

or

$$R(\theta_1, \theta_2, \theta_3) = \sum_{mm'} \left[\sum_l c_{lmm'} d^l_{m'm}(\theta_2) \right] \exp[i(m'\theta_3 + m\theta_1)].$$

The coefficients $c_{lmm'}$ refer to a particular pair of Patterson densities and are independent of the rotation. The coefficients $D^l_{m'm}$, containing the whole rotational part, refer to rotations of spherical harmonics and are independent of the particular Patterson densities. Since the summations over m and m' represent a Fourier synthesis, rapid calculation is possible.

As polar coordinates rather than Eulerian angles provide a more graphic interpretation of the rotation function, Tanaka (1977) has recast the initial definition as

$$\begin{aligned} R(\theta_1, \theta_2, \theta_3) = & \int_{\text{sphere}} [\mathscr{R}(\theta_1, \theta_2, \theta_3 = 0) P_1(r, \psi, \varphi)] \\ & \times [\mathscr{R}(\theta_1, \theta_2, \theta_3) P_2(r, \psi, \varphi)] \, \mathrm{d}V \\ = & \int_{\text{sphere}} [P_1(r, \psi, \varphi)][\mathscr{R}^{-1}(\theta_1, \theta_2, \theta_3 = 0) \\ & \times \mathscr{R}(\theta_1, \theta_2, \theta_3) P_2(r, \psi, \varphi)] \, \mathrm{d}V. \end{aligned}$$

He showed that the polar coordinates are now equivalent to $\kappa = \theta_3$, $\psi = \theta_2$ and $\varphi = \theta_1 - \pi/2$. The rotation function can then be expressed as

$$\begin{aligned} R(\kappa, \psi, \varphi) = \sum_{lmm'} \left(\sum_n a^*_{lmn} b_{lm'n} \right) \sum_q \{ & d^l_{qm}(\psi) d^l_{qm'}(\psi) (-1)^{(m'-m)} \\ & \times \exp[i(\kappa q)] \exp[i(m' - m)\varphi] \}, \end{aligned}$$

permitting rapid calculation of the fast rotation function in polar coordinates.

Crowther (1972) uses the Eulerian angles α, β, γ which are related to those defined by Rossmann & Blow (1962) according to $\theta_1 = \alpha + \pi/2$, $\theta_2 = \beta$ and $\theta_3 = \gamma - \pi/2$.

An alternative formulation of the fast rotation function, which reduces the errors in the calculation, is implemented in *AMoRe* (Navaza, 1987, 1993, 1994, 2001*a*). New target functions derived from the principle of maximum likelihood have been implemented in conjunction with fast rotation functions in the program *Phaser*, which can also take advantage of partial model information in orienting unknown fragments (Storoni *et al.*, 2004).

2.3.6.6. Locked rotation functions

Many oligomers of macromolecules obey simple point-group symmetry, which is maintained as noncrystallographic symmetry when they are crystallized. For example, a homo-tetramer often obeys 222 point-group symmetry, and icosahedral viruses obey 532 point-group symmetry. The locked rotation function takes advantage of this information and can greatly simplify the calculation and the interpretation of rotation functions (Fig. 2.3.6.6) (Arnold *et al.*, 1984; Rossmann *et al.*, 1972; Tong, 2001*a*; Tong & Rossmann, 1990, 1997). During the rotation-function calculation, the noncrystallographic symmetry of the crystal is *locked* to the presumed point group, hence the name locked rotation function.

Given the noncrystallographic symmetry point group, a standard orientation can be defined which serves as a reference orientation for this point group. For example, for 222 point-group symmetry, the standard orientation can be defined such that the three twofold axes are parallel to the three Cartesian coordinate axes that are defined with respect to the crystal unit cell. Once the standard orientation is defined, any orientation of the noncrystallographic symmetry point group can be related to the standard

orientation by a single set of three rotation angles that determine the rotation matrix [**E**].

Assume $[\mathbf{I}_n]$ ($n = 1, \ldots, N$) is the collection of noncrystallographic symmetry point-group rotation matrices in the standard orientation. Then the operation [**E**] will bring the noncrystallographic symmetry point group to a new orientation and the noncrystallographic symmetry rotation matrices in this new orientation, $[\boldsymbol{\rho}_n]$, are given by (Tong & Rossmann, 1990)

$$[\boldsymbol{\rho}_n] = [\mathbf{E}][\mathbf{I}_n][\mathbf{E}]^{-1}. \qquad (2.3.6.8)$$

For each rotation [**E**], the ordinary self-rotation-function value (R_n) for each of the noncrystallographic symmetry rotation matrices in the new orientation ($[\boldsymbol{\rho}_n]$) is calculated. The locked self-rotation-function value ($R_{\rm L}$) for this rotation is defined as the average of the ordinary rotation-function values over the noncrystallographic symmetry elements

$$R_{\rm L}([\mathbf{E}]) = \frac{1}{N-1}\sum_{n=2}^{N} R_n,$$

where the summation starts from 2 as it is assumed that $[\mathbf{I}_1]$ is the identity matrix.

The locked self rotation function simplifies the task of interpreting the self rotation function for the orientation of an noncrystallographic symmetry assembly. Instead of searching for $N-1$ peaks in the ordinary self rotation function, a single peak is sought in the locked self rotation function. It must be emphasized that this rotation ([**E**]) in the locked self rotation function is most often a general rotation. The locked self rotation function also reduces the noise in the rotation-function calculation by a factor of $(N-1)^{1/2}$ due to the averaging of the ordinary rotation-function values (Tong & Rossmann, 1990).

The symmetry of the locked self rotation function is generally rather complex and an analytical solution is often impossible (Tong & Rossmann, 1990). It depends not only on the crystallographic symmetry and the noncrystallographic symmetry, but also on the definition of the standard orientation of the noncrystallographic symmetry. For example, if the standard orientation is defined such that the twofold axes are parallel to the Cartesian coordinate axes for the 222 point group, a 90° rotation around the X, Y or Z axis, or a 120° rotation around the 111 direction, does not cause a net change to the standard orientation. Such rotations will appear as symmetry in the locked self rotation function (Tong & Rossmann, 1997). In practice, the locked self rotation function can be calculated rather quickly, especially if the fast rotation function is used. A large region of rotation space can be explored in the calculation of the locked rotation function and the solutions can then be clustered based on the resulting orientation of the noncrystallographic symmetry. For example, two rotations $[\mathbf{E}_1]$ and $[\mathbf{E}_2]$ that produce the same set of noncrystallographic symmetry matrices based on (2.3.6.8) are likely to be related by the symmetry of the locked self rotation function.

A locked cross rotation function can also be defined to determine the orientation, [**F**], of the known monomer structure relative to the noncrystallographic symmetry of the molecular assembly (Navaza *et al.*, 1998; Tong, 2001*a*; Tong & Rossmann, 1990, 1997). With the knowledge of [**F**] and the orientation of the noncrystallographic symmetry in the crystal [**E**], which can be determined from the locked self rotation function, the orientation of all the monomers in the crystal cell is given by

$$[\boldsymbol{\rho}_n] = [\mathbf{E}][\mathbf{I}_n][\mathbf{F}].$$

Therefore, $[\boldsymbol{\rho}_n]$ represents the rotational relationship between the monomer search model and the monomers of the assembly in the crystal. An ordinary cross-rotation-function value R_n can be calculated for each of the rotations $[\boldsymbol{\rho}_n]$, and the locked cross-rotation-function value is defined as the average

$$R_{\rm L} = (1/N)\sum_n R_n.$$

Like the locked self rotation function, the locked cross rotation function can determine the orientation of all the monomers of the noncrystallographic symmetry assembly with a single rotation.

2.3.7. Translation functions

2.3.7.1. Introduction

The problem of determining the position of a noncrystallographic symmetry element in space, or the position of a molecule of known orientation in a unit cell, has been reviewed by Rossmann (1972), Colman *et al.* (1976), Karle (1976), Argos & Rossmann (1980), Harada *et al.* (1981) and Beurskens (1981). All methods depend on the prior knowledge of the object's orientation implied by the rotation matrix [***C***]. The various translation functions, T, derived below, can only be computed given this information.

The general translation function can be defined as

$$T(\mathbf{S}_x, \mathbf{S}_{x'}) = \int_U \rho_1(\mathbf{x}) \cdot \rho_2(\mathbf{x}')\, {\rm d}\mathbf{x},$$

where T is a six-variable function given by each of the three components that define $\mathbf{S}_x$ and $\mathbf{S}_{x'}$. Here $\mathbf{S}_x$ and $\mathbf{S}_{x'}$ are equivalent reference positions of the objects, whose densities are $\rho_1(\mathbf{x})$ and $\rho_2(\mathbf{x}')$. The translation function searches for the optimal overlap of the two objects after they have been similarly oriented. Following the same procedure used for the rotation-function derivation, Fourier summations are substituted for $\rho_1(\mathbf{x})$ and $\rho_2(\mathbf{x}')$. It can then be shown that

$$T(\mathbf{S}_x, \mathbf{S}_{x'}) = \int_U \left\{\frac{1}{V_{\mathbf{h}}}\sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}| \exp[i(\alpha_{\mathbf{h}} - 2\pi\mathbf{h}\cdot\mathbf{x})]\right\} \times \left\{\frac{1}{V_{\mathbf{p}}}\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{p}}| \exp[i(\alpha_{\mathbf{p}} - 2\pi\mathbf{p}\cdot\mathbf{x}')]\right\} {\rm d}\mathbf{x}.$$

Using the substitution $\mathbf{x}' = [\boldsymbol{C}]\mathbf{x} + \mathbf{d}$ and simplifying leads to

$$T(\mathbf{S}_x, \mathbf{S}'_x) = \frac{1}{V_{\mathbf{h}}V_{\mathbf{p}}}\sum_{\mathbf{h}}\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| \times \exp[i(\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi\mathbf{p}\cdot\mathbf{d})] \times \int_U \exp\{-2\pi i(\mathbf{h} + [\boldsymbol{C}]^T\mathbf{p})\cdot\mathbf{x}\}\, {\rm d}\mathbf{x}.$$

The integral is the diffraction function $G_{\mathbf{hp}}$ (2.3.6.4). If the integration is taken over the volume U, centred at $\mathbf{S}_x$ and $\mathbf{S}_{x'}$, it follows that

$$T(\mathbf{S}_x, \mathbf{S}_{x'}) = \frac{2}{V_{\mathbf{h}}V_{\mathbf{p}}}\sum_{\mathbf{h}}\sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| G_{\mathbf{hp}} \times \cos[\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi(\mathbf{h}\cdot\mathbf{S}_x + \mathbf{p}\cdot\mathbf{S}_{x'})]. \qquad (2.3.7.1)$$

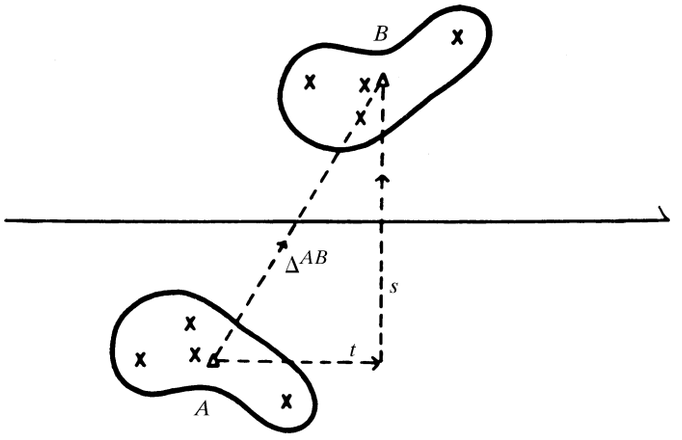


Fig. 2.3.7.1. Crosses represent atoms in a two-dimensional model structure. The triangles are the points chosen as approximate centres of molecules A and B. Δ^{AB} has components t and s parallel and perpendicular, respectively, to the screw rotation axis. [Reprinted from Rossmann *et al.* (1964).]

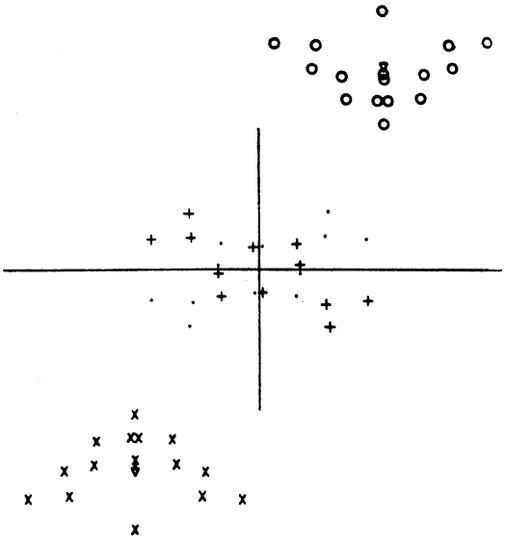


Fig. 2.3.7.2. Vectors arising from the structure in Fig. 2.3.7.1. The self-vectors of molecules A and B are represented by + and ·; the cross-vectors from molecules A to B and B to A by × and ○. Triangles mark the position of $+\Delta^{AB}$ and $-\Delta^{AB}$. [Reprinted from Rossmann *et al.* (1964).]

2.3.7.2. Position of a noncrystallographic element relating two unknown structures

The function (2.3.7.1) is quite general. For instance, the rotation function corresponds to a comparison of Patterson functions P_1 and P_2 at their origins. That is, the coefficients are F^2, phases are zero and $\mathbf{S}_x = \mathbf{S}_{x'} = 0$. However, the determination of the translation between two objects requires the comparison of cross-vectors away from the origin.

Consider, for instance, the determination of the precise translation vector parallel to a rotation axis between two identical molecules of unknown structure. For simplicity, let the noncrystallographic axis be a dyad (Fig. 2.3.7.1). Fig. 2.3.7.2 shows the corresponding Patterson of the hypothetical point-atom structure. Opposite sets of cross-Patterson vectors in Fig. 2.3.7.2 are related by a twofold rotation and a translation equal to twice the precise vector in the original structure. A suitable translation function would then compare a Patterson at $\mathbf{S}$ with the rotated Patterson at $-\mathbf{S}$. Hence, substituting $\mathbf{S}_x = \mathbf{S}$ and $\mathbf{S}_{x'} = -\mathbf{S}$ in (2.3.7.1),

$$T(\mathbf{S}) = \frac{2}{V^2}\sum_{\mathbf{h}}\sum_{\mathbf{p}}|\mathbf{F}_\mathbf{h}|^2|\mathbf{F}_\mathbf{p}|^2 G_{\mathbf{hp}}\cos[2\pi(\mathbf{h}-\mathbf{p})\cdot\mathbf{S}]. \qquad (2.3.7.2)$$

The opposite cross-vectors can be superimposed only if an evenfold rotation between the unknown molecules exists. The translation function (2.3.7.2) is thus applicable only in this special situation. There is no published translation method to determine the interrelation of two unknown structures in a crystallographic asymmetric unit or in two different crystal forms. However, another special situation exists if a molecular evenfold axis is parallel to a crystallographic evenfold axis. In this case, the position of the noncrystallographic symmetry element can be easily determined from the large peak in the corresponding Harker section of the Patterson.

In general, it is difficult or impossible to determine the positions of noncrystallographic axes (or their intersection at a molecular centre). However, the position of heavy atoms in isomorphous derivatives, which usually obey the noncrystallographic symmetry, can often determine this information.

2.3.7.3. Position of a known molecular structure in an unknown unit cell

The most common type of translation function occurs when looking for the position of a known molecular structure in an unknown crystal. For instance, if the structure of an enzyme has previously been determined by the isomorphous replacement method, then the structure of the same enzyme from another species can often be solved by molecular replacement [*e.g.* Grau *et al.* (1981)]. However, there are some severe pitfalls when, for instance, there are gross conformational changes [*e.g.* Moras *et al.* (1980)]. This type of translation function could also be useful in the interpolation of E maps produced by direct methods. Here there may often be confusion as a consequence of a number of molecular images related by translations (Karle, 1976; Beurskens, 1981; Egert & Sheldrick, 1985).

Tollin's (1966) Q function and Crowther & Blow's (1967) translation function are essentially identical (Tollin, 1969) and depend on a prior knowledge of the search molecule as well as its orientation in the unknown cell. The derivation given here, however, is somewhat more general and follows the derivation of Argos & Rossmann (1980), and should be compared with the method of Harada *et al.* (1981).

If the known molecular structure is correctly oriented into a cell (**p**) of an unknown structure and placed at **S** with respect to a defined origin, then a suitable translation function is

$$T(\mathbf{S}) = \sum_{\mathbf{p}}|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2|\mathbf{F}_\mathbf{p}(\mathbf{S})|^2. \qquad (2.3.7.3)$$

This definition is preferable to one based on an R-factor calculation as it is more amenable to computation and is independent of a relative scale factor.

The structure factor $\mathbf{F}_\mathbf{p}(\mathbf{S})$ can be calculated by modifying expression (2.3.8.9) (see below). That is,

$$\mathbf{F}_\mathbf{p}(\mathbf{S}) = \frac{U}{V_\mathbf{h}}\sum_{n=1}^{N}\exp(2\pi i\mathbf{p}\cdot\mathbf{S}_n)\left[\sum_{\mathbf{h}}\mathbf{F}_\mathbf{h}G_{\mathbf{hp}_n}\exp(-2\pi i\mathbf{h}\cdot\mathbf{S})\right],$$

where $V_\mathbf{h}$ is the volume of cell (**h**) and $\mathbf{S}_n$ is the position, in the nth crystallographic asymmetric unit, of cell (**p**) corresponding to **S** in known cell (**h**). Let

$$A_{\mathbf{p},\,n}\exp(i\gamma_n) = \sum_{\mathbf{h}}\mathbf{F}_\mathbf{h}G_{\mathbf{hp}_n}\exp(-2\pi i\mathbf{h}\cdot\mathbf{S}),$$

which are the coefficients of the molecular transform for the known molecule placed into the nth asymmetric unit of the **p** cell. Thus

$$\mathbf{F}_{\mathbf{p}}(\mathbf{S}) = \frac{U}{V_{\mathbf{h}}} \sum_{n=1}^{N} A_{\mathbf{p},n} \exp[i(\gamma + 2\pi \mathbf{p} \cdot \mathbf{S}_n)]$$

or

$$\mathbf{F}_{\mathbf{p}}(\mathbf{S}) = \frac{U}{V_{\mathbf{h}}} \sum_{n=1}^{N} A_{\mathbf{p},n} \exp[i(\gamma_n + 2\pi \mathbf{p}_n \cdot \mathbf{S})],$$

where $\mathbf{p}_n = [\boldsymbol{C}_n^T]\mathbf{p}$ and $\mathbf{S} = \mathbf{S}_1$. Hence

$$|\mathbf{F}_{\mathbf{p}}(\mathbf{S})|^2 = \left(\frac{U}{V_{\mathbf{h}}}\right)^2 \sum_n \sum_m \big(A_{\mathbf{p},n} A_{\mathbf{p},m} \times \exp\{i[2\pi(\mathbf{p}_n - \mathbf{p}_m) \cdot \mathbf{S} + (\gamma_n - \gamma_m)]\}\big),$$

and then from (2.3.7.3)

$$T(\mathbf{S}) = \left(\frac{U}{V_{\mathbf{h}}}\right)^2 \sum_{\mathbf{p}} \sum_n \sum_m \Big(|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2 A_{\mathbf{p},n} A_{\mathbf{p},m} \times \exp\{i[2\pi(\mathbf{p}_n - \mathbf{p}_m) \cdot \mathbf{S} + (\gamma_n - \gamma_m)]\}\Big), \qquad (2.3.7.4)$$

which is a Fourier summation with known coefficients $\{|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2 A_{\mathbf{p},n} A_{\mathbf{p},m} \times \exp[i(\gamma_n - \gamma_m)]\}$ such that $T(\mathbf{S})$ will be a maximum at the correct molecular position.

Terms with $n = m$ in expression (2.3.7.4) can be omitted as they are independent of **S** and only contribute a constant to the value of $T(\mathbf{S})$. For terms with $n \neq m$, the indices take on special values. For instance, if the **p** cell is monoclinic with its unique axis parallel to **b** such that $\mathbf{p}_1 = (p, q, r)$ and $\mathbf{p}_2 = (\bar{p}, q, \bar{r})$, then $\mathbf{p}_1 - \mathbf{p}_2$ would be $(2p, 0, 2r)$. Hence, $T(\mathbf{S})$ would be a two-dimensional function consistent with the physical requirement that the translation component, parallel to the twofold monoclinic axis, is arbitrary.

Crowther & Blow (1967) show that if $\mathbf{F}_{\mathrm{M}}$ are the structure factors of a known molecule correctly oriented within the cell of the unknown structure at an arbitrary molecular origin, then (altering the notation very slightly from above)

$$T(\mathbf{S}) = \sum_{\mathbf{p}} |\mathbf{F}_{\mathrm{obs}}(\mathbf{p})|^2 \mathbf{F}_{\mathrm{M}}(\mathbf{p}) \mathbf{F}_{\mathrm{M}}^*(\mathbf{p}[\boldsymbol{C}]) \exp(-2\pi i \mathbf{p} \cdot \mathbf{S}),$$

where $[\boldsymbol{C}]$ is a crystallographic symmetry operator relative to which the molecular origin is to be determined. This is of the same form as (2.3.7.4) but concerns the special case where the **h** cell, into which the known molecule was placed, has the same dimensions as the **p** cell.

The translation function as defined by (2.3.7.4) is on an arbitrary scale, which makes it difficult to compare results from different calculations. Translation functions can also be defined based on the crystallographic R factor or a correlation coefficient (CC). In particular, CCs based on reflection intensities can be evaluated by Fourier methods (Navaza & Vernoslova, 1995), although it is still computationally more expensive than the evaluation of (2.3.7.4). Alternatively, the translation function can be calculated first with (2.3.7.4), and then the R factor and CC can be calculated for the resulting top solutions.

A correct solution should also produce satisfactory packing arrangements of the molecular models in the crystal. Packing functions have been derived that estimate the amount of overlap among the models (Harada *et al.*, 1981; Hendrickson & Ward, 1976; Rabinovich & Shakked, 1984; Simpson *et al.*, 2001), and such considerations can frequently limit the search volume very considerably. Alternatively, a simple enumeration of the actual close contacts among different molecules in the crystal (for example, Cα–Cα distances less than 3 Å) has also been found to be an effective way of eliminating those solutions that produce unreasonable crystal packing (Jogl *et al.*, 2001; Tong, 1993). If conformational differences are expected between the search atomic model and the actual structure, care must be taken when applying this packing check.

2.3.7.4. Position of a noncrystallographic symmetry element in a poorly defined electron-density map

If an initial set of poor phases, for example from an SIR derivative, are available and the rotation function has given the orientation of a noncrystallographic rotation axis, it is possible to search the electron-density map systematically to determine the translation axis position. The translation function must, therefore, measure the quality of superposition of the poor electron-density map on itself. Hence $\mathbf{S}_x = \mathbf{S}_{x'} = \mathbf{S}$ and the function (2.3.7.1) now becomes

$$T(\mathbf{S}) = \frac{2}{V_{\mathbf{h}}^2} \sum_{\mathbf{h}} \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| G_{\mathbf{hp}} \cos[\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi(\mathbf{h} + \mathbf{p}) \cdot \mathbf{S}].$$

This real-space translation function has been used successfully to determine the intermolecular dyad axis for α-chymotrypsin (Blow *et al.*, 1964) and to verify the position of immunoglobulin domains (Colman & Fehlhammer, 1976).

2.3.7.5. Locked translation function

In a translation search, an atomic model with a given orientation is moved systematically through the unit cell. In such a situation, the structure-factor equation takes on the special form (Harada *et al.*, 1981; Rae, 1977; Tong, 1993)

$$\mathbf{F}_{\mathbf{h}}^{\mathrm{c}} = \sum_n \mathbf{F}_{\mathbf{h},n} \exp(2\pi i \mathbf{h}^T [\mathbf{T}_n] \mathbf{S}),$$

where **S** is the translation vector and the summation goes over the crystallographic symmetry operators. $\mathbf{F}_{\mathbf{h},n}$ is the structure factor calculated based only on the nth symmetry-related molecule,

$$\mathbf{F}_{\mathbf{h},n} = \sum_j f_j \exp\{2\pi i \mathbf{h}^T ([\mathbf{T}_n] \mathbf{x}_j^0 + \mathbf{t}_n)\},$$

where $\mathbf{x}_j^0$ represents the atomic position of the model at the reference position and the summation goes over all the atoms.

Noting equation (2.3.7.3), the translation function is given by

$$T(\mathbf{S}) = \sum_{\mathbf{h}} \sum_n |\mathbf{F}_{\mathbf{h}}^{\mathrm{o}}|^2 |\mathbf{F}_{\mathbf{h},n}|^2 + \sum_{\mathbf{h}} \sum_n \sum_{m \neq n} |\mathbf{F}_{\mathbf{h}}^{\mathrm{o}}|^2 |\mathbf{F}_{\mathbf{h},n} \mathbf{F}_{\mathbf{h},m}^*| \exp\{-2\pi i \mathbf{h}^T ([\mathbf{T}_m] - [\mathbf{T}_n]) \mathbf{S}\}, \qquad (2.3.7.5)$$

where the second term is the ordinary translation function, analogous to (2.3.7.4). The first term of (2.3.7.5) depends on the orientation of the model. Maximization of this term, or its correlation coefficient equivalent, is the basis behind the Patterson-correlation refinement (Brünger, 1990; Tong, 1996*b*) and the direct rotation function (DeLano & Brünger, 1995). It is

also related to the intensity-based domain refinement (Yeates & Rini, 1990).

In the presence of noncrystallographic symmetry, the locked self rotation function can be used to define the orientation of the noncrystallographic symmetry point group in the crystal. If an atomic model is available for the monomer but not for the entire oligomer, the locked cross rotation function can be used to determine the orientation of this monomer in the oligomer. The locked translation function can then be used to determine the position of this monomer relative to the centre of the noncrystallographic symmetry point group (Tong, 1996*b*, 2001*a*), which will produce a model for the entire oligomer. The centre of this oligomer in the crystal can be defined by a simple translation search.

With the knowledge of the orientation of one monomer of the oligomer, the first term of (2.3.7.5) is dependent on the position of this monomer relative to the centre of the noncrystallographic symmetry oligomer (Tong, 1996*b*). The atomic positions of the entire noncrystallographic symmetry oligomer in the standard orientation are given by

$$\mathbf{X}_{n,j} = [\mathbf{I}_n]\big([\mathbf{F}]\mathbf{X}_j^0 + \mathbf{V}_0\big),$$

where $\mathbf{X}_j^0$ are the atomic positions of the monomer model, centred at (0, 0, 0); [**F**] is the orientation of this model in the oligomer in the standard orientation; $\mathbf{V}_0$ is the position of this monomer relative to the centre of the oligomer; and $[\mathbf{I}_n]$ is the nth noncrystallographic symmetry rotation matrix in the standard orientation. The atomic positions of the noncrystallographic symmetry oligomer in the crystal unit cell, centred at the origin, are given by

$$\mathbf{x}_{n,j} = [\boldsymbol{\alpha}][\mathbf{E}]\mathbf{X}_{n,j} = [\boldsymbol{\alpha}][\mathbf{E}][\mathbf{I}_n]\big([\mathbf{F}]\mathbf{X}_j^0 + \mathbf{V}_0\big),$$

where [**E**] is the orientation of the noncrystallographic symmetry in the crystal unit cell and [$\boldsymbol{\alpha}$] is the deorthogonalization matrix.

By incorporating the calculated structure factors based on this noncrystallographic symmetry oligomer into the first term of (2.3.7.5), the locked translation function is given by

$$\begin{aligned} T_{\rm L}(\mathbf{V}_0) &= \sum_{\mathbf{h}} |\mathbf{F}_{\mathbf{h}}^{\rm o}|^2 |\mathbf{F}_{\mathbf{h}}|^2 \\ &= \sum_{\mathbf{h}} \sum_{n} \sum_{m \neq n} |\mathbf{F}_{\mathbf{h}}^{\rm o}|^2 \mathbf{F}_{\mathbf{h},n} \mathbf{F}_{\mathbf{h},m}^{*} \exp\{-2\pi i \mathbf{h}([\boldsymbol{\theta}_m] - [\boldsymbol{\theta}_n])\mathbf{V}_0\}, \end{aligned} \tag{2.3.7.6}$$

where $[\boldsymbol{\theta}_n] = [\boldsymbol{\alpha}][\mathbf{E}][\mathbf{I}_n]$ and $\mathbf{F}_{\mathbf{h},n_o} = \sum_j f_j \exp(2\pi i \mathbf{h}[\boldsymbol{\theta}_n][\mathbf{F}]\mathbf{X}_j^0)$. A constant term $\sum_{\mathbf{h}} \sum_n |\mathbf{F}_{\mathbf{h}}^{\rm o}|^2 |\mathbf{F}_{\mathbf{h},n}|^2$ has been omitted from this equation.

Conceptually, the locked translation function is based on the overlap of intermolecular vectors within the noncrystallographic symmetry oligomer and the observed Patterson map (Tong, 1996*b*). The equation for the locked translation function, (2.3.7.6), bears remarkable resemblance to that for the ordinary Patterson-correlation translation function, (2.3.7.5), with the interchange of the crystallographic ([$\mathbf{T}_n$]) and noncrystallographic symmetry ([$\boldsymbol{\theta}_n$]) parameters.

2.3.7.6. Computer programs for rotation and translation function calculations

Several programs are currently in popular use for the calculation of rotation and translation functions. These include *AMoRe* (Navaza, 1994, 2001*a*), *BEAST* (Read, 2001*b*), *CCP4* (Collaborative Computational Project, Number 4, 1994), *CNS* (Brünger *et al.*, 1998), *COMO* (Jogl *et al.*, 2001), *EPMR* (Kissinger *et al.*, 1999), *GLRF* (part of the *Replace* package) (Tong, 1993, 2001*a*; Tong & Rossmann, 1990, 1997), *Molrep* (Vagin & Teplyakov, 2000) and *Phaser* (Storoni *et al.*, 2004).

The correct placement of an atomic model in a crystal unit cell is generally a six-dimensional problem, with three degrees of rotational freedom and three degrees of translational freedom. Systematic examination of all six degrees of freedom at the same time is computationally expensive and cannot be used routinely (Fujinaga & Read, 1987; Rabinovich & Shakked, 1984; Sheriff *et al.*, 1999). On the other hand, directed sampling of the six degrees of freedom, driven by a stochastic or genetic algorithm (Chang & Lewis, 1997; Glykos & Kokkinidis, 2000; Kissinger *et al.*, 1999), has been successful in solving structures.

Traditionally, the calculations are divided into a rotational component (the rotation function) and a translational component (the translation function). Only a few rotation angles (for example the top few peaks of the rotation function) are manually passed to the translation function for examination (Fitzgerald, 1988). With the power of modern computers, it is now possible to perform limited six-dimensional searches, with the sampling of the rotational degrees of freedom guided by the rotation function. For example, the top peaks of the rotation function (Navaza, 1994) and their neighbours (Urzhumtsev & Podjarny, 1995) can be automatically examined by the translation function. A more general approach is to examine all rotation-function grid points with values greater than a certain threshold (Tong, 1996*a*). Such combined molecular replacement protocols have been found to be very powerful in solving new structures.

2.3.8. Molecular replacement

2.3.8.1. Using a known molecular fragment

The most straightforward application of the molecular replacement method occurs when the orientation and position of a known molecular fragment in an unknown cell have been previously determined. The simple procedure is to apply the rotation and translation operations to the known fragment. This will place it into one 'standard' asymmetric unit of the unknown cell. Then the crystal operators (assuming no further noncrystallographic operators are present in the unknown cell) are applied to generate the complete unit cell of the unknown structure. Structure factors can then be calculated from the rotated and translated known molecule into the unknown cell. The resultant model can be refined in numerous ways.

More generally, consider a molecule placed in any crystal cell (**h**), within which coordinate positions shall be designated by **x**. Let the corresponding structure factors be $\mathbf{F}_{\mathbf{h}}$. It is then possible to compute the structure factors $\mathbf{F}_{\mathbf{p}}$ for another cell (**p**) into which the same molecule has been placed N times related by the crystallographic symmetry operators $[\boldsymbol{C}_1], \mathbf{d}_1; [\boldsymbol{C}_2], \mathbf{d}_2; \ldots; [\boldsymbol{C}_N], \mathbf{d}_N$. Let the electron density at a point $\mathbf{y}_1$ in the first crystallographic asymmetric unit be spatially related to the point $\mathbf{y}_n$ in the nth asymmetric unit of the **p** crystal such that

$$\rho(\mathbf{y}_n) = \rho(\mathbf{y}_1), \tag{2.3.8.1}$$

where

$$\mathbf{y}_n = [\boldsymbol{C}_n]\mathbf{y}_1 + \mathbf{d}_n. \tag{2.3.8.2}$$

From the definition of a structure factor,

$$\mathbf{F}_{\mathbf{p}} = \sum_{n=1}^{N} \int_U \rho(\mathbf{y}_n) \exp(2\pi i \mathbf{p} \cdot \mathbf{y}_n)\, {\rm d}\mathbf{y}_n, \tag{2.3.8.3}$$

where the integral is taken over the volume U of one molecule. But since each molecule is identical as expressed in equation (2.3.8.1) and since (2.3.8.2) can be substituted in equation (2.3.8.3), we have

$$\mathbf{F}_\mathbf{p} = \sum_{n=1}^{N} \int_U \rho(\mathbf{y}_1) \exp[2\pi i \mathbf{p} \cdot ([\boldsymbol{C}_n]\mathbf{y}_1 + \mathbf{d}_n)] \, \mathrm{d}\mathbf{y}_1. \quad (2.3.8.4)$$

Now let the molecule in the **h** crystal be related to the molecule in the first asymmetric unit of the **p** crystal by the noncrystallographic symmetry operation

$$\mathbf{x} = [\boldsymbol{C}]\mathbf{y} + \mathbf{d}, \quad (2.3.8.5)$$

which implies

$$\rho(\mathbf{x}) = \rho(\mathbf{y}_1) = \rho(\mathbf{y}_2) = \ldots. \quad (2.3.8.6)$$

Furthermore, in the **h** cell

$$\rho(\mathbf{x}) = \frac{1}{V_\mathbf{h}} \sum_\mathbf{h} \mathbf{F}_\mathbf{h} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}), \quad (2.3.8.7)$$

and thus, by combining with (2.3.8.5), (2.3.8.6) and (2.3.8.7),

$$\rho(\mathbf{y}_1) = \frac{1}{V_\mathbf{h}} \sum_\mathbf{h} \mathbf{F}_\mathbf{h} \exp[-2\pi i(\mathbf{h}[\boldsymbol{C}] \cdot \mathbf{y}_1 + \mathbf{h} \cdot \mathbf{d})]. \quad (2.3.8.8)$$

Now using (2.3.8.4) and (2.3.8.8) it can be shown that

$$\mathbf{F}_\mathbf{p} = \frac{U}{V_\mathbf{h}} \sum_\mathbf{h} \mathbf{F}_\mathbf{h} \sum_{n=1}^{N} G_{\mathbf{hp}_n} \exp[2\pi i(\mathbf{p} \cdot \mathbf{S}_n - \mathbf{h} \cdot \mathbf{S})], \quad (2.3.8.9)$$

where

$$UG_{\mathbf{hp}_n} = \int_U \exp[2\pi i(p[\boldsymbol{C}_n] - \mathbf{h}[\boldsymbol{C}]) \cdot \mathbf{u}] \, \mathrm{d}\mathbf{u}. \quad (2.3.8.10)$$

S is a chosen molecular origin in the **h** crystal and $\mathbf{S}_n$ is the corresponding molecular position in the nth asymmetric unit of the **p** crystal.

2.3.8.2. Using noncrystallographic symmetry for phase improvement

The use of noncrystallographic symmetry for phase determination was proposed by Rossmann & Blow (1962, 1963) and subsequently explored by Crowther (1967, 1969) and Main & Rossmann (1966). These methods were developed in reciprocal space and were primarily concerned with *ab initio* phase determination. Real-space averaging of electron density between noncrystallographically related molecules was used in the structure determination of deoxyhaemoglobin (Muirhead *et al*., 1967) and of α-chymotrypsin (Matthews *et al*., 1967). The improvement derived from the averaging between the two noncrystallographic units was, however, not clear in either case. The first obviously successful application was in the structure determination of lobster glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al*., 1974; Argos *et al*., 1975), where the tetrameric molecule of symmetry 222 occupied one crystallographic asymmetric unit. The improvement in the essentially SIR electron-density map was considerable and the results changed from uninterpretable to interpretable. The uniqueness and validity of the solution lay in the obvious chemical correctness of the polypeptide fold and its agreement with known amino-acid-sequence data. In contrast to the earlier reciprocal-space methods, noncrystallographic symmetry was used as a method to improve poor phases rather than to determine phases *ab initio*.

Table 2.3.8.1. *Molecular replacement: phase refinement as an iterative process*

(*A*)	$\mathbf{F}_\mathrm{obs}, \alpha'_n, m'_n \rightarrow \rho_n$
(*B*)	$\rho_n \rightarrow \rho_n$(modified) (i) Use of noncrystallographic symmetry operators (ii) Definition of envelope limiting volume within which noncrystallographic symmetry is valid (iii) Adjustment of solvent density† (iv) Use of crystallographic operators to reconstruct modified density into a complete cell
(*C*)	ρ_n(modified) $\rightarrow \mathbf{F}_{\mathrm{calc},\,n+1}; \alpha_{\mathrm{calc},\,n+1}$
(*D*)	$(\mathbf{F}_{\mathrm{calc},\,n+1}, \alpha_{\mathrm{calc},\,n+1}) + (\mathbf{F}_\mathrm{obs}, \alpha_0) \rightarrow \mathbf{F}_\mathrm{obs}, \alpha'_{n+1}, m'_{n+1}$ (i) Assessment of reliability of new phasing set α_{n+1} in relation to original phasing set $\alpha_0(w)$ (ii) Use of figures of merit m_0, m_{n+1} and reliability w to determine modified phasing set α'_{n+1}, m'_{n+1}‡ (iii) Consideration of α_{n+1} and m_{n+1} where there was no prior knowledge of (*a*) $\mathbf{F}_\mathrm{obs}$ (*e.g.* very low order reflections or uncollected data) (*b*) α_0 (*e.g.* no isomorphous information or phase extension)
(*E*)	Return to step (*A*) with α'_{n+1}, m'_{n+1} and a possibly augmented set of $\mathbf{F}_\mathrm{obs}$.

† Wang (1985); Bhat & Blow (1982); Collins (1975); Schevitz *et al*. (1981); Hoppe & Gassmann (1968). ‡ Rossmann & Blow (1961); Hendrickson & Lattman (1970).

Many other applications followed rapidly, aided greatly by the versatile techniques developed by Bricogne (1976). Of particular interest is the application to the structure determination of hexokinase (Fletterick & Steitz, 1976), where the averaging occurred both between different crystal forms and within the same crystal.

The most widely used procedure for real-space averaging is the 'double sorting' technique developed by Bricogne (1976) and also by Johnson (1978). An alternative method is to maintain the complete map stored in the computer (Nordman, 1980*b*). This avoids the sorting operation, but is only possible given a very large computer or a low-resolution map containing relatively few grid points.

Bricogne's double sorting technique involves generating real-space non-integral points (D_i) which are related to integral grid points (I_i) in the cell asymmetric unit by the noncrystallographic symmetry operators. The elements of the set D_i are then brought back to their equivalent points in the cell asymmetric unit (D'_i) and sorted by their proximity to two adjacent real-space sections. The set I'_i, calculated on a finer grid than I_i and stored in the computer memory two sections at a time, is then used for linear interpolation to determine the density values at D'_i which are successively stored and summed in the related array I_i. A count is kept of the number of densities received at each I_i, resulting in a final averaged aggregate, when all real-space sections have been utilized. The density to be assigned outside the molecular envelope (defined with respect to the set I_i) is determined by averaging the density of all unused points in I_i. The grid interval for the set I'_i should be about one-sixth of the resolution to avoid serious errors from interpolation (Bricogne, 1976). The grid point separation in the set I_i need only be sufficient for representation of electron density, or about one-third of the resolution.

Molecular replacement in real space consists of the following steps (Table 2.3.8.1): (*a*) calculation of electron density based on a starting phase set and observed amplitudes; (*b*) averaging of this density among the noncrystallographic asymmetric units or molecular copies in several crystal forms, a process which defines

a molecular envelope as the averaging is only valid within the range of the noncrystallographic symmetry; (*c*) reconstructing the unit cell based on averaged density in every noncrystallographic asymmetric unit; (*d*) calculating structure factors from the reconstructed cell; (*e*) combining the new phases with others to obtain a weighted best-phase set; and (*f*) returning to step (*a*) at the previous or an extended resolution. Decisions made in steps (*b*) and (*e*) determine the rate of convergence (see Table 2.3.8.1) to a solution (Arnold *et al*., 1987).

The power of the molecular replacement procedure for either phase improvement or phase extension depends on the number of noncrystallographic asymmetric units, the size of the excluded volume expressed in terms of the ratio $(V - UN)/V$ and the magnitude of the measurement error on the structure amplitudes. Crowther (1967, 1969) and Bricogne (1974) have investigated the dependence on the number of noncrystallographic asymmetric units and conclude that three or more copies are sufficient to ensure convergence of an iterative phase improvement procedure in the absence of errors on the structure amplitudes. As with the analogous case of isomorphous replacement in which three data sets ensure reasonable phase determination, additional copies will enhance the power of the method, although their usefulness is subject to the law of diminishing returns. Another example of this principle is the sign determination of the *h*0*l* reflections of horse haemoglobin (Perutz, 1954) in which seven shrinkage stages constituted the sampling of the transform of a single copy.

In an analysis of how phasing errors propagate into errors in calculations of electron density, Arnold & Rossmann (1986) concluded that the 'power' of phase determination could be related to the noncrystallographic redundancy, *N*, the ratio of the molecular envelope volume, *U*, to the unit cell volume, *V*, the fractional error of the structure-factor amplitudes, *R* and the fractional completeness of the data, *f*, by (Arnold & Rossmann, 1986)

$$P = \frac{(Nf)^{1/2}}{RU/V}. \qquad (2.3.8.11)$$

This semiquantitative result makes intuitive sense in that the noncrystallographic redundancy and solvent content terms can be directly related to over-sampling of the molecular transform in reciprocal space, and, thus, are analogous in providing phasing information. The phasing power of solvent flattening/density modification was further analysed and shown to lead to Sayre's equations (Sayre, 1952) at a limit where the molecular envelope is sufficiently detailed and shrunken to cover sharpened and separated atoms (Arnold & Rossmann, 1986). This result suggests that more detailed definitions of molecular envelopes than are traditionally used could be advantageous for phase improvement and extension procedures.

Procedures for real-space averaging have been used extensively with great success. The interesting work of Wilson *et al*. (1981) is noteworthy for the continuous adjustment of molecular envelope with increased map definition. Furthermore, the analysis of complete virus structures has only been possible as a consequence of this technique (Bloomer *et al*., 1978; Harrison *et al*., 1978; Abad-Zapatero *et al*., 1980; Liljas *et al*., 1982). Although the procedure has been used primarily for phase improvement, apparently successful attempts have been made at phase extension (Nordman, 1980*b*; Gaykema *et al*., 1984; Rossmann *et al*., 1985). *Ab initio* phasing of glyceraldehyde-3-phosphate dehydrogenase (Argos *et al*., 1975) was successfully attempted by initially filling the known envelope with uniform density to determine the phases of the innermost reflections and then gradually extending phases to 6.3 Å resolution. Johnson *et al*. (1976) used the same procedure to determine the structure of southern bean mosaic virus to 22.5 Å resolution. Particularly impressive was the work on polyoma virus (Rayment *et al*., 1982; Rayment, 1983; Rayment *et al*., 1983) where crude initial models led to an entirely unexpected breakdown of the Caspar & Klug (1962) concept of quasi-symmetry. *Ab initio* phasing has also been used by combining the electron-diffraction projection data of two different crystal forms of bacterial rhodopsin (Rossmann & Henderson, 1982).

2.3.8.3. Update on noncrystallographic averaging and density-modification methods

Since this article was originally written, molecular replacement has been subject of a number of reviews (Rossmann, 1990), including a historical background of the subject (Rossmann, 2001). A series of chapters pertaining to molecular replacement have been published in *IT* Volume F (Rossmann & Arnold, 2001*a*), reviewing noncrystallographic symmetry (Chapter 13.1; Blow, 2001), rotation (Chapter 13.2; Navaza, 2001*b*) and translation (Chapter 13.3; Tong, 2001*b*) functions, and noncrystallographic symmetry averaging for phase improvement and extension (Chapter 13.4; Rossmann & Arnold, 2001*b*). Chapters on phase improvement by density modification (Chapter 15.1; Zhang *et al.*, 2001), optimal weighting of Fourier terms in map calculations (Chapter 15.2; Read, 2001*a*) and refinement calculations incorporating bulk solvent correction (Chapter 18.4; Dauter *et al.*, 2001) are also recommended reading.

There has been remarkable progress in the general area of density modification, involving improvement of real-space methods for averaging and reconstruction, and treatment of solvent for iterative phase improvement and refinement calculations. The use of real-space averaging between noncrystallographically related electron density within the crystallographic asymmetric unit has become an accepted mode of extending phase information to higher resolution, particularly for complex structures such as viruses [Acharya *et al.*, 1989; Arnold & Rossmann, 1988; Gaykema *et al.*, 1986; Hogle *et al.*, 1985; Luo *et al.*, 1989; Rossmann & Arnold, 2001*b* (*IT* F Chapter 13.4); Rossmann *et al.*, 1985, 1992]. *Ab initio* phase determination based on noncrystallographic redundancy has become fairly common (Chapman *et al.*, 1992; Lunin *et al.*, 2000; Miller *et al.*, 2001; Rossmann, 1990; Tsao *et al.*, 1992). General programs in common use for noncrystallographic symmetry averaging include *BUSTER-TNT* [Blanc *et al.*, 2004; Roversi *et al.*, 2000; Tronrud & Ten Eyck, 2001 (*IT* F Section 25.2.4)], *CNS* [Brünger *et al.*, 1998; Brunger, Adams, DeLano *et al.*, 2001 (*IT* F Section 25.2.3)], *DM*/*DMMULTI* [Cowtan & Main, 1993; Cowtan *et al.*, 2001 (*IT* F Section 25.2.2); Schuller, 1996; Zhang, 1993], *PHASES* [Furey, 2001 (*IT* F Section 25.2.1); Furey & Swaminathan, 1997], *RAVE*/*MAVE* (Jones, 1992; Kleywegt, 1996) and *SOLVE*/*RESOLVE* [Terwilliger, 2002*b*, 2003*c*; Terwilliger & Berendzen, 2001 (*IT* F Section 14.2.2)].

Solvent flattening has been formulated in reciprocal space for greater computational efficiency (Leslie, 1987; Terwilliger, 1999) and solvent 'flipping' is a powerful extension of solvent density modification (Abrahams, 1997; Abrahams & Leslie, 1996). Bulk-solvent corrections are now commonly used in crystallographic refinement, allowing for better modelling and phase determination of low-resolution data [Brünger *et al.*, 1998; Dauter *et al.*, 2001 (*IT* F Chapter 18.4)]. The problem of phase error estimation and analysis and bias removal has been treated extensively (Cowtan, 1999; Cowtan & Main, 1996), including extension of methods to include maximum-likelihood functions and iterative bias removal procedures [Brunger, Adams & Rice, 2001 (*IT* F Chapter 18.2); Hunt & Deisenhofer, 2003; Lamzin *et al.*, 2001 (*IT* F Section 25.2.5); Perrakis *et al.*, 1997; Terwilliger, 2004]. Histogram matching [Cowtan & Main, 1993; Lunin, 1993; Nieh & Zhang, 1999; Refaat *et al.*, 1996; Zhang, 1993; Zhang *et al.*, 2001 (*IT* F Chapter 15.1)] and skeletonization [Baker *et al.*, 1993;

Zhang *et al.*, 2001 (*IT* F Chapter 15.1)], and structural fragment matching procedures (Terwilliger, 2003*a*) have been added to the arsenal of density-modification methods. Automated mask and molecular-envelope definition has helped to remove the tedium and increase the efficiency and quality of density-modification and symmetry-averaging procedures. Noncrystallographic symmetry averaging among different crystal forms (Perutz, 1954) has become increasingly common, and exploitation of the unit-cell variation among flash-cooled and noncooled forms of the same crystal is a broadly applicable method for phase determination (Das *et al.*, 1996; Ding *et al.*, 1995); soaking crystals in a series of different solvents and buffers can produce an analogous effect (Ren *et al.*, 1995; Tong *et al.*, 1997). Phases from noncrystallographic symmetry averaging and other 'experimental' sources have been incorporated into crystallographic refinement procedures using a number of formalisms (Arnold & Rossmann, 1988; Rees & Lewis, 1983) including maximum likelihood (Pannu *et al.*, 1998).

2.3.8.4. Equivalence of real- and reciprocal-space molecular replacement

Let us proceed in reciprocal space doing exactly the same as is done in real-space averaging. Thus

$$\rho_{\rm AV}(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\rho(\mathbf{x}_n),$$

where

$$\mathbf{x}_n = [\boldsymbol{C}_n]\mathbf{x} + \mathbf{d}_n.$$

Therefore,

$$\rho_{\rm AV}(\mathbf{x}) = \frac{1}{N}\sum_{N}\frac{1}{V}\left[\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_n)\right].$$

The next step is to perform the back-transform of the averaged electron density. Hence,

$$\mathbf{F}_{\mathbf{p}} = \int_U \rho_{\rm AV}(\mathbf{x})\exp(-2\pi i\mathbf{p}\cdot\mathbf{x})\,{\rm d}\mathbf{x},$$

where U is the volume within the averaged part of the cell. Hence, substituting for $\rho_{\rm AV}$,

$$\mathbf{F}_{\mathbf{p}} = \int_U\left[\frac{1}{NV}\sum_{N}\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_n)\right]\exp(-2\pi i\mathbf{p}\cdot\mathbf{x})\,{\rm d}\mathbf{x},$$

which is readily simplified to

$$\mathbf{F}_{\mathbf{p}} = \frac{U}{NV}\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\sum_{N}G_{\mathbf{hp}n}\exp(2\pi i\mathbf{h}\cdot\mathbf{d}_n).$$

Setting

$$\mathbf{B}_{\mathbf{hp}} = \frac{U}{NV}\sum_{N}G_{\mathbf{hp}n}\exp(2\pi i\mathbf{h}\cdot\mathbf{d}_n),$$

the molecular replacement equations can be written as

$$\mathbf{F}_{\mathbf{p}} = \sum_{\mathbf{h}}\mathbf{B}_{\mathbf{hp}}\mathbf{F}_{\mathbf{h}} \qquad (2.3.8.12)$$

(Main & Rossmann, 1966), or in matrix form

$$\mathbf{F} = [\boldsymbol{B}]\mathbf{F},$$

which is the form of the equations used by Main (1967) and by Crowther (1967). Colman (1974) arrived at the same conclusions by an application of Shannon's sampling theorem. It should be noted that the elements of $[\boldsymbol{B}]$ are dependent only on knowledge of the noncrystallographic symmetry and the volume within which it is valid. Substitution of approximate phases into the right-hand side of (2.3.8.12) produces a set of calculated structure factors exactly analogous to those produced by back-transforming the averaged electron density in real space. The new phases can then be used in a renewed cycle of molecular replacement. The reciprocal-space molecular replacement procedure has been implemented and tested in a computer program (Tong & Rossmann, 1995).

Computationally, it has been found more convenient and faster to work in real space. This may, however, change with the advent of vector processing in 'supercomputers'. Obtaining improved phases by substitution of current phases on the right-hand side of the molecular replacement equations (2.3.8.1) seems less cumbersome than the repeated forward and backward Fourier transformation, intermediate sorting, and averaging required in the real-space procedure.

2.3.9. Conclusions

Complete interpretation of Patterson maps is no longer used frequently in structure analysis, although most determinations of heavy-atom positions of isomorphous pairs are based on Patterson analyses. Incorporation of the Patterson concept is crucial in many sophisticated techniques essential for the solution of complex problems, particularly in the application to biological macromolecular structures. Patterson techniques provide important physical insights in a link between real- and reciprocal-space formulation of crystal structures and diffraction data.

This article, first written in December 1984 (by MGR and EA) and completed in January 1986, was published in the first edition of this volume 1993, and in a mildly revised form in the second edition in 2001. We are grateful for generous support of our laboratories from the National Science Foundation (to LT and MGR) and from the National Institutes of Health (LT, MGR and EA). We acknowledge the many authors whose insights, innovation and writings make up the subject matter of this review. We also acknowledge Sharon Wilder for her painstaking attention to detail in preparation of the original manuscript and an article by Argos & Rossmann (1980) as the source of some material in this article.

References

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Structure of southern bean mosaic virus at 2.8 Å resolution. Nature (London)*, **286**, 33–39.

Abrahams, J. P. (1997). *Bias reduction in phase refinement by modified interference functions: introducing the γ correction. Acta Cryst.* D**53**, 371–376.

Abrahams, J. P. & Leslie, A. G. W. (1996). *Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Cryst.* D**52**, 30–42.

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown, F. (1989). *The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. Nature (London)*, **337**, 709–716.

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. & Sheat, S. (1969). *Structure of rhombohedral 2 zinc insulin crystals. Nature (London)*, **224**, 491–495.

Åkervall, K., Strandberg, B., Rossmann, M. G., Bengtsson, U., Fridborg, K., Johannisen, H., Kannan, K. K., Lövgren, S., Petef, G., Öberg, B., Eaker, D., Hjertén, S., Rydén, L. & Moring, I. (1972). *X-ray diffraction studies of the structure of satellite tobacco necrosis virus. Cold Spring Harbor Symp. Quant. Biol.* **36**, 469–488.

Argos, P., Ford, G. C. & Rossmann, M. G. (1975). *An application of the molecular replacement technique in direct space to a known protein structure. Acta Cryst.* A**31**, 499–506.

Argos, P. & Rossmann, M. G. (1974). *Determining heavy-atom positions using non-crystallographic symmetry. Acta Cryst.* A**30**, 672–677.

Argos, P. & Rossmann, M. G. (1976). *A method to determine heavy-atom positions for virus structures. Acta Cryst.* B**32**, 2975–2979.

Argos, P. & Rossmann, M. G. (1980). *Molecular replacement methods.* In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 361–417. New York: Plenum.

Arnold, E., Erickson, J. W., Fout, G. S., Frankenberger, E. A., Hecht, H. J., Luo, M., Rossmann, M. G. & Rueckert, R. R. (1984). *Virion orientation in cubic crystals of the human common cold virus HRV14. J. Mol. Biol.* **177**, 417–430.

Arnold, E. & Rossmann, M. G. (1986). *Effect of errors, redundancy, and solvent content in the molecular replacement procedure for the structure determination of biological macromolecules. Proc. Natl Acad. Sci. USA*, **83**, 5489–5493.

Arnold, E. & Rossmann, M. G. (1988). *The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. Acta Cryst.* A**44**, 270–283.

Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). *The structure determination of a common cold virus, human rhinovirus 14. Acta Cryst.* A**43**, 346–361.

Baker, D., Bystroff, C., Fletterick, R. J. & Agard, D. A. (1993). *PRISM: topologically constrained phase refinement for macromolecular crystallography. Acta Cryst.* D**49**, 429–439.

Beevers, C. A. & Robertson, J. M. (1950). *Interpretation of the Patterson synthesis. Acta Cryst.* **3**, 164.

Beurskens, P. T. (1981). *A statistical interpretation of rotation and translation functions in reciprocal space. Acta Cryst.* A**37**, 426–430.

Bhat, T. N. & Blow, D. M. (1982). *A density-modification method for the improvement of poorly resolved protein electron-density maps. Acta Cryst.* A**38**, 21–29.

Bijvoet, J. M. (1954). *Structure of optically active compounds in the solid state. Nature (London)*, **173**, 888–891.

Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. (1951). *Determination of the absolute configuration of optically active compounds by means of X-rays. Nature (London)*, **168**, 271–272.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Cryst.* D**60**, 2210–2221.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Blow, D. M. (1958). *The structure of haemoglobin. VII. Determination of phase angles in the noncentrosymmetric [100] zone. Proc. R. Soc. London Ser. A*, **247**, 302–336.

Blow, D. M. (2001). *Noncrystallographic symmetry*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.1. Dordrecht: Kluwer Academic Publishers.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Blow, D. M. & Rossmann, M. G. (1961). *The single isomorphous replacement method. Acta Cryst.* **14**, 1195–1202.

Blow, D. M., Rossmann, M. G. & Jeffery, B. A. (1964). *The arrangement of α-chymotrypsin molecules in the monoclinic crystal form. J. Mol. Biol.* **8**, 65–78.

Bluhm, M. M., Bodo, G., Dintzis, H. M. & Kendrew, J. C. (1958). *The crystal structure of myoglobin. IV. A Fourier projection of sperm-whale myoglobin by the method of isomorphous replacement. Proc. R. Soc. London Ser. A*, **246**, 369–389.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography.* New York: Academic Press.

Bodo, G., Dintzis, H. M., Kendrew, J. C. & Wyckoff, H. W. (1959). *The crystal structure of myoglobin. V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals. Proc. R. Soc. London Ser. A*, **253**, 70–102.

Bragg, W. L. (1958). *The determination of the coordinates of heavy atoms in protein crystals. Acta Cryst.* **11**, 70–75.

Bragg, W. L. & Perutz, M. F. (1954). *The structure of haemoglobin. VI. Fourier projections on the 010 plane. Proc. R. Soc. London Ser. A*, **225**, 315–329.

Braun, P. B., Hornstra, J. & Leenhouts, J. I. (1969). *Automated crystal-structure determination by Patterson search using a known part of the molecule. Philips Res. Rep.* **24**, 85–118.

Bricogne, G. (1974). *Geometric sources of redundancy in intensity data and their use for phase determination. Acta Cryst.* A**30**, 395–405.

Bricogne, G. (1976). *Methods and programs for the direct space exploitation of geometric redundancies. Acta Cryst.* A**32**, 832–847.

Brünger, A. T. (1990). *Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. Acta Cryst.* A**46**, 46–57.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Cryst.* D**54**, 905–921.

Brunger, A. T., Adams, P. D., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Pannu, N. S., Read, R. J., Rice, L. M. & Simonson, T. (2001). *The structure-determination language of the Crystallography & NMR System*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.3. Dordrecht: Kluwer Academic Publishers.

Brunger, A. T., Adams, P. D. & Rice, L. M. (2001). *Enhanced macromolecular refinement by simulated annealing*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 18.2. Dordrecht: Kluwer Academic Publishers.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *Structure determination of crystalline lobster* D*-glyceraldehyde-3-phosphate dehydrogenase. J. Mol. Biol.* **82**, 563–585.

Buerger, M. J. (1946). *The interpretation of Harker syntheses. J. Appl. Phys.* **17**, 579–595.

Buerger, M. J. (1950*a*). *Some new functions of interest in X-ray crystallography. Proc. Natl Acad. Sci. USA*, **36**, 376–382.

Buerger, M. J. (1950*b*). *Limitation of electron density by the Patterson function. Proc. Natl Acad. Sci. USA*, **36**, 738–742.

Buerger, M. J. (1951). *A new approach to crystal-structure analysis. Acta Cryst.* **4**, 531–544.

Buerger, M. J. (1953*a*). *Image theory of superposed vector sets. Proc. Natl Acad. Sci. USA*, **39**, 669–673.

Buerger, M. J. (1953*b*). *Solution functions for solving superposed Patterson syntheses. Proc. Natl Acad. Sci. USA*, **39**, 674–678.

Buerger, M. J. (1953*c*). *An intersection function and its relations to the minimum function of X-ray crystallography. Proc. Natl Acad. Sci. USA*, **39**, 678–680.

Buerger, M. J. (1959). *Vector Space and its Application in Crystal-Structure Investigation.* New York: John Wiley.

Buerger, M. J. (1966). *Background for the use of image-seeking functions. Trans. Am. Crystallogr. Assoc.* **2**, 1–9.

Bullough, R. K. (1961). *On homometric sets. I. Some general theorems. Acta Cryst.* **14**, 257–269.

Bullough, R. K. (1964). *On homometric sets. II. Sets obtained by singular transformations. Acta Cryst.* **17**, 295–308.

Burdina, V. I. (1970). *Symmetry of the rotation function. Kristallografiya*, **15**, 623–630.

Burdina, V. I. (1971). *Symmetry of the rotation function. Sov. Phys. Crystallogr.* **15**, 545–550.

Burdina, V. I. (1973). *Primitive rotation regions of two Patterson syntheses. Kristallografiya*, **18**, 694–700.

Burnett, R. M. & Rossmann, M. G. (1971). *The determination of the crystal structure of trans-2,4-dihydroxy-2,4-dimethylcyclohexane-trans-*

1-acetic acid γ-lactone, $C_{10}H_{16}O_3$, using rotation and translation functions in reciprocal space. Acta Cryst. B**27**, 1378–1387.
Carlisle, C. H. & Crowfoot, D. (1945). *The crystal structure of cholesteryl iodide. Proc. R. Soc. London Ser. A*, **184**, 64–83.
Caspar, D. L. D. & Klug, A. (1962). *Physical principles in the construction of regular viruses. Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.
Chang, G. & Lewis, M. (1997). *Molecular replacement using genetic algorithms. Acta Cryst.* D**53**, 279–289.
Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Ab initio phase determination for spherical viruses: parameter determination for spherical-shell models*. *Acta Cryst.* A**48**, 301–312.
Clastre, J. & Gay, R. (1950). *La détermination des structures cristallines à partir du diagramme de Patterson. Compt. Rend.* **230**, 1876–1877.
Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography*. *Acta Cryst.* D**50**, 760–763.
Collins, D. M. (1975). *Efficiency in Fourier phase refinement for protein crystal structures. Acta Cryst.* A**31**, 388–389.
Colman, P. M. (1974). *Noncrystallographic symmetry and the sampling theorem. Z. Kristallogr.* **140**, 344–349.
Colman, P. M. & Fehlhammer, H. (1976). *Appendix: the use of rotation and translation functions in the interpretation of low resolution electron density maps. J. Mol. Biol.* **100**, 278–282.
Colman, P. M., Fehlhammer, H. & Bartels, K. (1976). *Patterson search methods in protein structure determination: β-trypsin and immunoglobulin fragments.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 248–258. Copenhagen: Munksgaard.
Corfield, P. W. R. & Rosenstein, R. D. (1966). *Maximum information from the minimum function. Trans. Am. Crystallogr. Assoc.* **2**, 17–28.
Cowtan, K. (1999). *Error estimation and bias correction in phase-improvement calculations*. *Acta Cryst.* D**55**, 1555–1567.
Cowtan, K. D. & Main, P. (1993). *Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Cryst.* D**49**, 148–157.
Cowtan, K. D. & Main, P. (1996). *Phase combination and cross validation in iterated density-modification calculations*. *Acta Cryst.* D**52**, 43–48.
Cowtan, K. D., Zhang, K. Y. J. & Main, P. (2001). *DM/DMMULTI software for phase improvement by density modification*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.2. Dordrecht: Kluwer Academic Publishers.
Crick, F. H. C. & Magdoff, B. S. (1956). *The theory of the method of isomorphous replacement for protein crystals. I. Acta Cryst.* **9**, 901–908.
Cromer, D. T. (1974). *Dispersion corrections for X-ray atomic scattering factors.* In *International Tables for X-ray Crystallography*, Vol. IV, edited by J. A. Ibers & W. C. Hamilton, pp. 148–151. Birmingham: Kynoch Press.
Crowther, R. A. (1967). *A linear analysis of the non-crystallographic symmetry problem. Acta Cryst.* **22**, 758–764.
Crowther, R. A. (1969). *The use of non-crystallographic symmetry for phase determination. Acta Cryst.* B**25**, 2571–2580.
Crowther, R. A. (1972). *The fast rotation function.* In *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.
Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). *The structure of haemoglobin. IX. A three-dimensional Fourier synthesis at 5.5 Å resolution: description of the structure. Proc. R. Soc. London Ser. A*, **265**, 161–187.
Das, K., Ding, J., Hsiou, Y., Clark, A. D. Jr, Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A. J., Boyer, P. L., Clark, P., Smith, R. H. Jr, Smith, M. B. K., Michejda, C. J., Hughes, S. H. & Arnold, E. (1996). *Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. J. Mol. Biol.* **264**, 1085–1100.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Novel approach to phasing proteins: derivatization by short cryo-soaking with halides*. *Acta Cryst.* D**56**, 232–237.
Dauter, Z., Murshudov, G. N. & Wilson, K. S. (2001). *Refinement at atomic resolution*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 18.4. Dordrecht: Kluwer Academic Publishers.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *In-house measurement of the sulfur anomalous signal and its use for phasing*. *Acta Cryst.* D**59**, 688–696.
DeLano, W. L. & Brünger, A. T. (1995). *The direct rotation function: rotational Patterson correlation search applied to molecular replacement*. *Acta Cryst.* D**51**, 740–748.
Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. Acta Cryst.* **14**, 1188–1195.
Dickerson, R. E., Kopka, M. L., Varnum, J. C. & Weinzierl, J. E. (1967). *Bias, feedback and reliability in isomorphous phase analysis. Acta Cryst.* **23**, 511–522.
Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. (1968). *A least-squares refinement method for isomorphous replacement. Acta Cryst.* B**24**, 997–1003.
Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H. & Arnold, E. (1995). *Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors*. *Nature Struct. Biol.* **2**, 407–415.
Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). *The crystal structure of insulin. III. Evidence for a 2-fold axis in rhombohedral zinc insulin. J. Mol. Biol.* **16**, 227–241.
Egert, E. (1983). *Patterson search – an alternative to direct methods. Acta Cryst.* A**39**, 936–940.
Egert, E. & Sheldrick, G. M. (1985). *Search for a fragment of known geometry by integrated Patterson and direct methods. Acta Cryst.* A**41**, 262–268.
Eisenberg, D. (1970). *X-ray crystallography and enzyme structure.* In *The Enzymes*, edited by P. D. Boyer, Vol. I, 3rd ed., pp. 1–89. New York: Academic Press.
Fitzgerald, P. M. D. (1988). *MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. J. Appl. Cryst.* **21**, 273–278.
Fletterick, R. J. & Steitz, T. A. (1976). *The combination of independent phase information obtained from separate protein structure determinations of yeast hexokinase. Acta Cryst.* A**32**, 125–132.
Fridrichsons, J. & Mathieson, A. McL. (1962). *Image-seeking. A brief study of its scope and comments on certain limitations. Acta Cryst.* **15**, 1065–1074.
Fujinaga, M. & Read, R. J. (1987). *Experiences with a new translation-function program. J. Appl. Cryst.* **20**, 517–521.
Fukuyama, K., Abdel-Meguid, S. S., Johnson, J. E. & Rossmann, M. G. (1983). *Structure of a T = 1 aggregate of alfalfa mosaic virus coat protein seen at 4.5 Å resolution. J. Mol. Biol.* **167**, 873–894.
Furey, W. (2001). *PHASES*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.1. Dordrecht: Kluwer Academic Publishers.
Furey, W. & Swaminathan, S. (1997). *PHASES-95: a program package for processing and analyzing diffraction data from macromolecules*. *Methods Enzymol.* **277**, 590–620.
Garrido, J. (1950*a*). *Sur la détermination des structures cristallines au moyen de la transformée de Patterson. Compt. Rend.* **230**, 1878–1879.
Garrido, J. (1950*b*). *Les coincidences fortuites dans la méthode des différences vectorielles. Compt. Rend.* **231**, 297–298.
Gaykema, W. P., Volbeda, A. & Hol, W. G. (1986). *Structure determination of Panulirus interruptus haemocyanin at 3.2 Å resolution. Successful phase extension by sixfold density averaging*. *J. Mol. Biol.* **187**, 255–275.
Gaykema, W. P. J., Hol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J. & Beintema, J. J. (1984). *3.2 Å structure of the copper-containing, oxygen-carrying protein Panulirus interruptus haemocyanin. Nature (London)*, **309**, 23–29.
Gibbs, J. W. (1898). *Remarks regarding Michelson's letter. Nature (London)*, **59**, 200.
Glykos, N. M. & Kokkinidis, M. (2000). *A stochastic approach to molecular replacement*. *Acta Cryst.* D**56**, 169–174.
Grau, U. M., Rossmann, M. G. & Trommer, W. E. (1981). *The crystallization and structure determination of an active ternary complex of pig heart lactate dehydrogenase. Acta Cryst.* B**37**, 2019–2026.
Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. Proc. R. Soc. London Ser. A*, **225**, 287–307.

Grosse-Kunstleve, R. W. & Brunger, A. T. (1999). *A highly automated heavy-atom search procedure for macromolecular structures*. *Acta Cryst.* D**55**, 1568–1577.

Hamilton, W. C. (1965). *The crystal structure of orthorhombic acetamide. Acta Cryst.* **18**, 866–870.

Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *A translation function combining packing and diffraction information: an application to lysozyme (high-temperature form). Acta Cryst.* A**37**, 398–406.

Harker, D. (1936). *The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. J. Chem. Phys.* **4**, 381–390.

Harker, D. (1956). *The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Cryst.* **9**, 1–9.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Tomato bushy stunt virus at 2.9 Å resolution. Nature (London)*, **276**, 368–373.

Hendrickson, W. A. (1991). *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation*. *Science*, **254**, 51–58.

Hendrickson, W. A. & Lattman, E. E. (1970). *Representation of phase probability distributions for simplified combination of independent phase information. Acta Cryst.* B**26**, 136–143.

Hendrickson, W. A. & Teeter, M. M. (1981). *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London)*, **290**, 107–113.

Hendrickson, W. A. & Ward, K. B. (1976). *A packing function for delimiting the allowable locations of crystallized macromolecules. Acta Cryst.* A**32**, 778–780.

High, D. F. & Kraut, J. (1966). *The crystal structure of androsterone. Acta Cryst.* **21**, 88–96.

Hirshfeld, F. L. (1968). *Symmetry in the generation of trial structures. Acta Cryst.* A**24**, 301–311.

Hodgkin, D. C., Kamper, J., Lindsey, J., MacKay, M., Pickworth, J., Robertson, J. H., Shoemaker, C. B., White, J. G., Prosen, R. J. & Trueblood, K. N. (1957). *The structure of vitamin B_{12}. I. An outline of the crystallographic investigation of vitamin B_{12}. Proc. R. Soc. London Ser. A*, **242**, 228–263.

Hogle, J. M., Chow, M. & Filman, D. J. (1985). *Three-dimensional structure of poliovirus at 2.9 Å resolution. Science*, **229**, 1358–1365.

Hoppe, W. (1957*a*). *Die Faltmolekülmethode und ihre anwendung in der Röntgenographischen Konstitutionsanalyse von Biflorin ($C_{20}H_{20}O_4$). Z. Elektrochem.* **61**, 1076–1083.

Hoppe, W. (1957*b*). *Die 'Faltmolekülmethode' – eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannter Molekülstruktur. Acta Cryst.* **10**, 750–751.

Hoppe, W. (1959). *Die Bestimmung genauer Schweratom-parameter in isomorphen azentrischen Kristallen. Acta Cryst.* **12**, 665–674.

Hoppe, W. (1962). *'Nahezu-Homometrische Lösungen' und Faltmolekülmethode. Z. Kristallogr.* **117**, 249–258.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Hosemann, R. & Bagchi, S. N. (1954). *On homometric structures. Acta Cryst.* **7**, 237–241.

Huber, R. (1965). *Die automatisierte Faltmolekülmethode. Acta Cryst.* **19**, 353–356.

Hughes, E. W. (1940). *The crystal structure of dicyandiamide. J. Am. Chem. Soc.* **62**, 1258–1267.

Hunt, J. F. & Deisenhofer, J. (2003). *Ping-pong cross-validation in real space: a method for increasing the phasing power of a partial model without risk of model bias*. *Acta Cryst.* D**59**, 214–224.

International Tables for Crystallography (2005). Vol. A, *Space-Group Symmetry*, edited by Th. Hahn. Heidelberg: Springer.

Jacobson, R. A., Wunderlich, J. A. & Lipscomb, W. N. (1961). *The crystal and molecular structure of cellobiose. Acta Cryst.* **14**, 598–607.

James, R. W. (1965). *The optical principles of the diffraction of X-rays.* Ithaca: Cornell University Press.

Jogl, G., Tao, X., Xu, Y. & Tong, L. (2001). *COMO: a program for combined molecular replacement*. *Acta Cryst.* D**57**, 1127–1134.

Johnson, J. E. (1978). *Appendix II. Averaging of electron density maps. Acta Cryst.* B**34**, 576–577.

Johnson, J. E., Akimoto, T., Suck, D., Rayment, I. & Rossmann, M. G. (1976). *The structure of southern bean mosaic virus at 22.5 Å resolution. Virology*, **75**, 394–400.

Johnson, J. E., Argos, P. & Rossmann, M. G. (1975). *Rotation function studies of southern bean mosaic virus at 22 Å resolution. Acta Cryst.* B**31**, 2577–2583.

Jones, T. A. (1992). *A, yaap, asap, @#*? A set of averaging programs*. In *Molecular Replacement*, edited by E. J. Dodson, S. Glover & W. Wolf, pp. 91–105. Warrington: SERC Daresbury Laboratory.

Karle, J. (1976). *Partial structures and use of the tangent formula and translation functions.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 155–164. Copenhagen: Munksgaard.

Karle, J. & Hauptman, H. (1964). *Positivity, point atoms, and Pattersons. Acta Cryst.* **17**, 392–396.

Kartha, G. (1961). *Isomorphous replacement method in non-centrosymmetric structures. Acta Cryst.* **14**, 680–686.

Kartha, G. & Parthasarathy, R. (1965). *Combination of multiple isomorphous replacement and anomalous dispersion data for protein structure determination. I. Determination of heavy-atom positions in protein derivatives. Acta Cryst.* **18**, 745–749.

Ketelaar, J. A. A. & de Vries, T. A. (1939). *The crystal structure of tetra phosphonitrile chloride, $P_4N_4Cl_8$. Recl Trav. Chim.* **58**, 1081–1099.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Rapid automated molecular replacement by evolutionary search*. *Acta Cryst.* D**55**, 484–491.

Kleywegt, G. J. (1996). *Use of non-crystallographic symmetry in protein structure refinement*. *Acta Cryst.* D**52**, 842–857.

Kraut, J. (1961). *The crystal structure of 2-amino-ethanol phosphate. Acta Cryst.* **14**, 1146–1152.

Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). *The ARP/wARP suite for automated construction and refinement of protein models*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.5. Dordrecht: Kluwer Academic Publishers.

Lattman, E. E. (1972). *Optimal sampling of the rotation function. Acta Cryst.* B**28**, 1065–1068.

Lattman, E. E. & Love, W. E. (1970). *A rotational search procedure for detecting a known molecule in a crystal. Acta Cryst.* B**26**, 1854–1857.

Lentz, P. J. Jr, Strandberg, B., Unge, T., Vaara, I., Borell, A., Fridborg, K. & Petef, G. (1976). *The determination of the heavy-atom substitution sites in the satellite tobacco necrosis virus. Acta Cryst.* B**32**, 2979–2983.

Leslie, A. G. W. (1987). *A reciprocal-space method for calculating a molecular envelope using the algorithm of B.C. Wang*. *Acta Cryst.* A**43**, 134–136.

Lifchitz, A. (1983). *On the choice of the model cell and the integration volume in the use of the rotation function. Acta Cryst.* A**39**, 130–139.

Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lövgren, S., Skoglund, U. & Strandberg, B. (1982). *Structure of satellite tobacco necrosis virus at 3.0 Å resolution. J. Mol. Biol.* **159**, 93–108.

Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures.* Ithaca: Cornell University Press.

Litvin, D. B. (1975). *The molecular replacement method. I. The rotation function problem, application to bovine liver catalase and STNV*. *Acta Cryst.* A**31**, 407–416.

Lu, G. (1999). *FINDNCS: a program to detect non-crystallographic symmetries in protein crystals from heavy-atom sites*. *J. Appl. Cryst.* **32**, 365–368.

Lunin, V. Y. (1993). *Electron-density histograms and the phase problem*. *Acta Cryst.* D**49**, 90–99.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Skovoroda, T. P., Urzhumtsev, A. G. & Podjarny, A. D. (2000). *Low-resolution ab initio phasing: problems and advances*. *Acta Cryst.* D**56**, 1223–1232.

Luo, M., Vriend, G., Kamer, G. & Rossmann, M. G. (1989). *Structure determination of Mengo virus*. *Acta Cryst.* B**45**, 85–92.

Luzzati, V. (1953). *Résolution d'une structure cristalline lorsque les positions d'une partie des atomes sont connues: traitement statistique. Acta Cryst.* **6**, 142–152.

McKenna, R., Xia, D., Willingmann, P., Ilag, L. L. & Rossmann, M. G. (1992). *Structure determination of the bacteriophage φX174*. *Acta Cryst.* B**48**, 499–511.

McLachlan, D. Jr & Harker, D. (1951). *Finding the signs of the F's from the shifted Patterson product. Proc. Natl Acad. Sci. USA*, **37**, 846–849.

Main, P. (1967). *Phase determination using non-crystallographic symmetry. Acta Cryst.* **23**, 50–54.

Main, P. & Rossmann, M. G. (1966). *Relationships among structure factors due to identical molecules in different crystallographic environments. Acta Cryst.* **21**, 67–72.

Matthews, B. W. (1966). *The determination of the position of anomalously scattering heavy atom groups in protein crystals. Acta Cryst.* **20**, 230–239.

Matthews, B. W. & Czerwinski, E. W. (1975). *Local scaling: a method to reduce systematic errors in isomorphous replacement and anomalous scattering measurements. Acta Cryst.* A**31**, 480–487.

Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). *Three-dimensional structure of tosyl-α-chymotrypsin. Nature (London),* **214**, 652–656.

Menzer, G. (1949). *Über die mehrdeutigkeit der Kristallstrukturbestimmung. Z. Naturforsch. Teil A*, **4**, 11–21.

Mighell, A. D. & Jacobson, R. A. (1963). *Analysis of three-dimensional Patterson maps using vector verification. Acta Cryst.* **16**, 443–445.

Miller, S. T., Hogle, J. M. & Filman, D. J. (2001). *Ab initio phasing of high-symmetry macromolecular complexes: successful phasing of authentic poliovirus data to 3.0 Å resolution. J. Mol. Biol.* **307**, 499–512.

Moncrief, J. W. & Lipscomb, W. N. (1966). *Structure of leurocristine methiodide dihydrate by anomalous scattering methods; relation to leurocristine (vincristine) and vincaleukoblastine (vinblastine). Acta Cryst.* **21**, 322–331.

Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P. & Giegé, R. (1980). *Crystal structure of yeast tRNA^{Asp}. Nature (London)*, **288**, 669–674.

Muirhead, H., Cox, J. M., Mazzarella, L. & Perutz, M. F. (1967). *Structure and function of haemoglobin. III. A three-dimensional Fourier synthesis of human deoxyhaemoglobin at 5.5 Å resolution. J. Mol. Biol.* **28**, 117–156.

Murthy, M. R. N., Reid, T. J. III, Sicignano, A., Tanaka, N. & Rossmann, M. G. (1981). *Structure of beef liver catalase. J. Mol. Biol.* **152**, 465–499.

Nagem, R. A. P., Polikarpov, I. & Dauter, Z. (2003). *Phasing on rapidly soaked ions. Methods Enzymol.* **374**, 120–137.

Navaza, J. (1987). *On the fast rotation function. Acta Cryst.* A**43**, 645–653.

Navaza, J. (1993). *On the computation of the fast rotation function. Acta Cryst.* D**49**, 588–591.

Navaza, J. (1994). *AMoRe: an automated package for molecular replacement. Acta Cryst.* A**50**, 157–163.

Navaza, J. (2001*a*). *Implementation of molecular replacement in AMoRe. Acta Cryst.* D**57**, 1367–1372.

Navaza, J. (2001*b*). *Rotation functions*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.2. Dordrecht: Kluwer Academic Publishers.

Navaza, J., Panepucci, E. H. & Martin, C. (1998). *On the use of strong Patterson function signals in many-body molecular replacement. Acta Cryst.* D**54**, 817–821.

Navaza, J. & Vernoslova, E. (1995). *On the fast translation functions for molecular replacement. Acta Cryst.* A**51**, 445–449.

Nieh, Y.-P. & Zhang, K. Y. J. (1999). *A two-dimensional histogram-matching method for protein phase refinement and extension. Acta Cryst.* D**55**, 1893–1900.

Nixon, P. E. (1978). *Overlapping Patterson peaks and direct methods: the structure of prostratin. Acta Cryst.* A**34**, 450–453.

Nordman, C. E. (1966). *Vector space search and refinement procedures. Trans. Am. Crystallogr. Assoc.* **2**, 29–38.

Nordman, C. E. (1972). *An application of vector space search methods to the Patterson function of myoglobin. Acta Cryst.* A**28**, 134–143.

Nordman, C. E. (1980*a*). *Vector-space Patterson search and other stored-function sampling procedures.* In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 5.01–5.13. Bangalore: Indian Academy of Sciences.

Nordman, C. E. (1980*b*). *Procedures for detection and idealization of non-crystallographic symmetry with application to phase refinement of the satellite tobacco necrosis virus structure. Acta Cryst.* A**36**, 747–754.

Nordman, C. E. & Nakatsu, K. (1963). *Interpretation of the Patterson function of crystals containing a known molecular fragment. The structure of an Alstonia alkaloid. J. Am. Chem. Soc.* **85**, 353–354.

Nordman, C. E. & Schilling, J. W. (1970). *Calculation and use of vector overlap weights in Patterson search and refinement.* In *Crystallographic Computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 110–114. Copenhagen: Munksgaard.

North, A. C. T. (1965). *The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. Acta Cryst.* **18**, 212–216.

Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *New method in X-ray crystal structure determination involving the use of anomalous dispersion. Phys. Rev.* **98**, 1857–1858.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Incorporation of prior phase information strengthens maximum-likelihood structure refinement. Acta Cryst.* D**54**, 1285–1294.

Patterson, A. L. (1934*a*). *A Fourier series representation of the average distribution of scattering power in crystals. Phys. Rev.* **45**, 763.

Patterson, A. L. (1934*b*). *A Fourier series method for the determination of the components of interatomic distances in crystals. Phys. Rev.* **46**, 372–376.

Patterson, A. L. (1935). *A direct method for the determination of the components of interatomic distances in crystals. Z. Kristallogr.* **90**, 517–542.

Patterson, A. L. (1939). *Homometric structures. Nature (London)*, **143**, 939–940.

Patterson, A. L. (1944). *Ambiguities in the X-ray analysis of crystal structures. Phys. Rev.* **65**, 195–201.

Patterson, A. L. (1949). *An alternative interpretation for vector maps. Acta Cryst.* **2**, 339–340.

Pauling, L. & Shappell, M. D. (1930). *The crystal structure of bixbyite and the C-modification of the sesquioxides. Z. Kristallogr.* **75**, 128–142.

Pepinsky, R., Okaya, Y. & Takeuchi, Y. (1957). *Theory and application of the $P_s(u)$ function and anomalous dispersion in direct determination of structures and absolute configuration in non-centric crystals. Acta Cryst.* **10**, 756.

Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Cryst.* D**53**, 448–455.

Perutz, M. F. (1954). *The structure of haemoglobin. III. Direct determination of the molecular transform. Proc. R. Soc. London Ser. A*, **225**, 264–286.

Perutz, M. F. (1956). *Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Cryst.* **9**, 867–873.

Phillips, D. C. (1966). *Advances in protein crystallography.* In *Advances in Structure Research by Diffraction Methods*, Vol. 2, edited by R. Brill & R. Mason, pp. 75–140. New York: John Wiley.

Poljak, R. J. (1963). *Heavy-atom attachment to crystalline lysozyme. J. Mol. Biol.* **6**, 244–246.

Rabinovich, D. & Shakked, Z. (1984). *A new approach to structure determination of large molecules by multi-dimensional search methods. Acta Cryst.* A**40**, 195–200.

Rae, A. D. (1977). *The use of structure factors to find the origin of an oriented molecular fragment. Acta Cryst.* A**33**, 423–425.

Ramachandran, G. N. & Raman, S. (1959). *Syntheses for the deconvolution of the Patterson function. Part I. General principles. Acta Cryst.* **12**, 957–964.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Phasing on anomalous signal of sulfurs: what is the limit? Acta Cryst.* D**59**, 1020–1027.

Raman, S. (1966). *Patterson functions and vector sets. Trans. Am. Crystallogr. Assoc.* **2**, 10–16.

Raman, S. & Lipscomb, W. N. (1961). *Two classes of functions for the location of heavy atoms and for solution of crystal structures. Z. Kristallogr.* **116**, 314–327.

Ramaseshan, S. & Abrahams, S. C. (1975). Editors. *Anomalous Scattering.* Copenhagen: Munksgaard.

Rao, S. N., Jih, J. H. & Hartsuck, J. A. (1980). *Rotation-function space groups. Acta Cryst.* A**36**, 878–884.

Rayment, I. (1983). *Molecular replacement method at low resolution: optimum strategy and intrinsic limitations as determined by calculations on icosahedral virus models. Acta Cryst.* A**39**, 102–116.

Rayment, I., Baker, T. S. & Caspar, D. L. D. (1983). *A description of the techniques and application of molecular replacement used to determine the structure of polyoma virus capsid at 22.5 Å resolution. Acta Cryst.* B**39**, 505–516.

Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. (1982). *Polyoma virus capsid structure at 22.5 Å resolution. Nature (London)*, **295**, 110–115.

Rayment, I., Johnson, J. E., Suck, D., Akimoto, T. & Rossmann, M. G. (1978). *An 11 Å resolution electron density map of southern bean mosaic virus. Acta Cryst.* B**34**, 567–578.

Read, R. J. (2001*a*). *Model phases: probabilities, bias and maps*. In *International Tables for Crystallography*, Vol. F, *Crystallography of*

Biological Macromolecules, edited by M. G. Rossmann & E. Arnold, ch. 15.2. Dordrecht: Kluwer Academic Publishers.

Read, R. J. (2001*b*). *Pushing the boundaries of molecular replacement with maximum likelihood*. *Acta Cryst.* D**57**, 1373–1382.

Rees, D. C. & Lewis, M. (1983). *Incorporation of experimental phases in a restrained least-squares refinement*. *Acta Cryst.* A**39**, 94–97.

Refaat, L. S., Tate, C. & Woolfson, M. M. (1996). *Direct-space methods in phase extension and phase refinement. IV. The double-histogram method*. *Acta Cryst.* D**52**, 252–256.

Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. & Stammers, D. (1995). *High resolution structures of HIV-1 RT from four RT-inhibitor complexes*. *Nature Struct. Biol.* **2**, 293–302.

Robertson, J. M. (1935). *An X-ray study of the structure of phthalocyanines. Part I. The metal-free, nickel, copper, and platinum compounds. J. Chem. Soc.* pp. 615–621.

Robertson, J. M. (1936). *An X-ray study of the phthalocyanines. Part II. Quantitative structure determination of the metal-free compound. J. Chem. Soc.* pp. 1195–1209.

Robertson, J. M. (1951). *Interpretation of the Patterson synthesis: rubidium benzyl penicillin. Acta Cryst.* **4**, 63–66.

Robertson, J. M. & Woodward, I. (1937). *An X-ray study of the phthalocyanines. Part III. Quantitative structure determination of nickel phthalocyanine. J. Chem. Soc.* pp. 219–230.

Rogers, D. (1951). *New methods of direct structure determination using modified Patterson maps. Research*, **4**, 295–296.

Rossmann, M. G. (1960). *The accurate determination of the position and shape of heavy-atom replacement groups in proteins. Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961*a*). *The position of anomalous scatterers in protein crystals. Acta Cryst.* **14**, 383–388.

Rossmann, M. G. (1961*b*). *Application of the Buerger minimum function to protein structures.* In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 252–265. Oxford: Pergamon Press.

Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.

Rossmann, M. G. (1990). *The molecular replacement method. Acta Cryst.* A**46**, 73–82.

Rossmann, M. G. (2001). *Molecular replacement – historical background*. *Acta Cryst.* D**57**, 1360–1366.

Rossmann, M. G. & Arnold, E. (2001*a*). Editors. *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*. Dordrecht: Kluwer Academic Publishers.

Rossmann, M. G. & Arnold, E. (2001*b*). *Noncrystallographic symmetry averaging of electron density for molecular-replacement phase refinement and extension*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.4. Dordrecht: Kluwer Academic Publishers.

Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). *Structure of a human common cold virus and functional relationship to other picornaviruses. Nature (London)*, **317**, 145–153.

Rossmann, M. G. & Blow, D. M. (1961). *The refinement of structures partially determined by the isomorphous replacement method. Acta Cryst.* **14**, 641–647.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst.* **15**, 24–31.

Rossmann, M. G. & Blow, D. M. (1963). *Determination of phases by the conditions of non-crystallographic symmetry. Acta Cryst.* **16**, 39–45.

Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). *The relative positions of independent molecules within the same asymmetric unit. Acta Cryst.* **17**, 338–342.

Rossmann, M. G., Ford, G. C., Watson, H. C. & Banaszak, L. J. (1972). *Molecular symmetry of glyceraldehyde-3-phosphate dehydrogenase. J. Mol. Biol.* **64**, 237–249.

Rossmann, M. G. & Henderson, R. (1982). *Phasing electron diffraction amplitudes with the molecular replacement method. Acta Cryst.* A**38**, 13–20.

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J.-B., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *Molecular replacement real-space averaging*. *J. Appl. Cryst.* **25**, 166–180.

Roversi, P., Blanc, E., Vonrhein, C., Evans, G. & Bricogne, G. (2000). *Modelling prior distributions of atoms for macromolecular refinement and completion*. *Acta Cryst.* D**56**, 1316–1323.

Sasada, Y. (1964). *The differential rotation function. Acta Cryst.* **17**, 611–612.

Sayre, D. (1952). *The squaring method: a new method for phase determination*. *Acta Cryst.* **5**, 60–65.

Schevitz, R. W., Podjarny, A. D., Zwick, M., Hughes, J. J. & Sigler, P. B. (1981). *Improving and extending the phases of medium- and low-resolution macromolecular structure factors by density modification. Acta Cryst.* A**37**, 669–677.

Schiltz, M., Fourme, R. & Prange, T. (2003). *Use of noble gases xenon and krypton as heavy atoms in protein structure determination*. *Methods Enzymol.* **374**, 83–119.

Schuller, D. J. (1996). *MAGICSQUASH: more versatile non-crystallographic averaging with multiple constraints*. *Acta Cryst.* D**52**, 425–434.

Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement problems*. *J. Appl. Cryst.* **32**, 98–101.

Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *The crystal structure of L_8-threonine. J. Am. Chem. Soc.* **72**, 2328–2349.

Sim, G. A. (1961). *Aspects of the heavy-atom method.* In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 227–235. Oxford: Pergamon Press.

Simonov, V. I. (1965). *Calculation of the phases of the structure amplitudes by Fourier transformation of the sum, product and minimum functions. Proc. Indian Acad. Sci.* A**62**, 213–223.

Simpson, A. A., Leiman, P. G., Tao, Y., He, Y., Badasso, M. O., Jardine, P. J., Anderson, D. L. & Rossmann, M. G. (2001). *Structure determination of the head-tail connector of bacteriophage φ29*. *Acta Cryst.* D**57**, 1260–1269.

Simpson, P. G., Dobrott, R. D. & Lipscomb, W. N. (1965). *The symmetry minimum function: high order image seeking functions in X-ray crystallography*. *Acta Cryst.* **18**, 169–179.

Singh, A. K. & Ramaseshan, S. (1966). *The determination of heavy atom positions in protein derivatives. Acta Cryst.* **21**, 279–280.

Smith, J. L., Hendrickson, W. A. & Addison, A. W. (1983). *Structure of trimeric haemerythrin. Nature (London)*, **303**, 86–88.

Speakman, J. C. (1949). *The crystal structures of the acid salts of some monobasic acids. Part I. Potassium hydrogen bisphenyl acetate. J. Chem. Soc.* pp. 3357–3365.

Stauffacher, C. V., Usha, R., Harrington, M., Schmidt, T., Hosur, M. V. & Johnson, J. E. (1987). *The structure of cowpea mosaic virus at 3.5 Å resolution.* In *Crystallography in Molecular Biology*, edited by D. Moras, J. Drenth, B. Strandberg, D. Suck & K. Wilson, pp. 293–308. New York, London: Plenum.

Steinrauf, L. K. (1963). *Two Fourier functions for use in protein crystallography*. *Acta Cryst.* **16**, 317–319.

Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Likelihood-enhanced fast rotation functions*. *Acta Cryst.* D**60**, 432–438.

Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination.* New York: Macmillan.

Strahs, G. & Kraut, J. (1968). *Low-resolution electron-density and anomalous-scattering-density maps of Chromatium high-potential iron protein. J. Mol. Biol.* **35**, 503–512.

Tanaka, N. (1977). *Representation of the fast-rotation function in a polar coordinate system. Acta Cryst.* A**33**, 191–193.

Taylor, W. J. (1953). *Fourier representation of Buerger's image-seeking minimum function. J. Appl. Phys.* **24**, 662–663.

Terwilliger, T. C. (1999). *Reciprocal-space solvent flattening*. *Acta Cryst.* D**55**, 1863–1871.

Terwilliger, T. C. (2002*a*). *Rapid automatic NCS identification using heavy-atom substructures*. *Acta Cryst.* D**58**, 2213–2215.

Terwilliger, T. C. (2002*b*). *Statistical density modification with non-crystallographic symmetry*. *Acta Cryst.* D**58**, 2082–2086.

Terwilliger, T. C. (2003*a*). *Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement*. *Acta Cryst.* D**59**, 1174–1182.

Terwilliger, T. C. (2003*b*). *SOLVE and RESOLVE: Automated structure solution and density modification*. *Methods Enzymol.* **374**, 22–37.

Terwilliger, T. C. (2003*c*). *Statistical density modification using local pattern matching*. *Acta Cryst.* D**59**, 1688–1701.

Terwilliger, T. C. (2004). *Using prime-and-switch phasing to reduce model bias in molecular replacement*. *Acta Cryst.* D**60**, 2144–2149.
Terwilliger, T. C. & Berendzen, J. (1999). *Automated MAD and MIR structure solution*. *Acta Cryst.* D**55**, 849–861.
Terwilliger, T. C. & Berendzen, J. (2001). *Automated MAD and MIR structure solution*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 14.2.2. Dordrecht: Kluwer Academic Publishers.
Terwilliger, T. C. & Eisenberg, D. (1983). *Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. Acta Cryst.* A**39**, 813–817.
Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). *Generalized method of determining heavy-atom positions using the difference Patterson function*. *Acta Cryst.* A**43**, 1–5.
Tollin, P. (1966). *On the determination of molecular location. Acta Cryst.* **21**, 613–614.
Tollin, P. (1969). *A comparison of the Q-functions and the translation function of Crowther and Blow*. *Acta Cryst.* A**25**, 376–377.
Tollin, P. & Cochran, W. (1964). *Patterson function interpretation for molecules containing planar groups. Acta Cryst.* **17**, 1322–1324.
Tollin, P., Main, P. & Rossmann, M. G. (1966). *The symmetry of the rotation function. Acta Cryst.* **20**, 404–407.
Tollin, P. & Rossmann, M. G. (1966). *A description of various rotation function programs. Acta Cryst.* **21**, 872–876.
Tong, L. (1993). *REPLACE, a suite of computer programs for molecular-replacement calculations*. *J. Appl. Cryst.* **26**, 748–751.
Tong, L. (1996*a*). *Combined molecular replacement*. *Acta Cryst.* A**52**, 782–784.
Tong, L. (1996*b*). *The locked translation function and other applications of a Patterson correlation function*. *Acta Cryst.* A**52**, 476–479.
Tong, L. (2001*a*). *How to take advantage of non-crystallographic symmetry in molecular replacement: 'locked' rotation and translation functions*. *Acta Cryst.* D**57**, 1383–1389.
Tong, L. (2001*b*). *Translation functions*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch. 13.3. Dordrecht: Kluwer Academic Publishers.
Tong, L., Qian, C., Davidson, W., Massariol, M.-J., Bonneau, P. R., Cordingley, M. G. & Lagacé, L. (1997). *Experiences from the structure determination of human cytomegalovirus protease*. *Acta Cryst.* D**53**, 682–690.
Tong, L. & Rossmann, M. G. (1990). *The locked rotation function. Acta Cryst.* A**46**, 783–792.
Tong, L. & Rossmann, M. G. (1993). *Patterson-map interpretation with noncrystallographic symmetry*. *J. Appl. Cryst.* **26**, 15–21.
Tong, L. & Rossmann, M. G. (1995). *Reciprocal-space molecular-replacement averaging*. *Acta Cryst.* D**51**, 347–353.
Tong, L. & Rossmann, M. G. (1997). *Rotation function calculations with GLRF program*. *Methods Enzymol.* **276**, 594–611.
Tronrud, D. E. & Ten Eyck, L. F. (2001). *The TNT refinement package*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, Section 25.2.4. Dordrecht: Kluwer Academic Publishers.
Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Ab initio phase determination for viruses with high symmetry: a feasibility study*. *Acta Cryst*. A**48**, 293–301.
Urzhumtsev, A. & Podjarny, A. (1995). *On the solution of the molecular-replacement problem at very low resolution: Application to large complexes*. *Acta Cryst.* D**51**, 888–895.
Vagin, A. & Teplyakov, A. (2000). *An approach to multi-copy search in molecular replacement*. *Acta Cryst.* D**56**, 1622–1624.
Wang, B. C. (1985). *Resolution of phase ambiguity in macromolecular crystallography*. *Methods Enzymol.* **115**, 90–112.
Weeks, C. M., Adams, P. D., Berendzen, J., Brunger, A. T., Dodson, E. J., Grosse-Kunstleve, R. W., Schneider, T. R., Sheldrick, G. M., Terwilliger, T. C., Turkenburg, M. G. & Uson, I. (2003). *Automatic solution of heavy-atom substructures*. *Methods Enzymol.* **374**, 37–83.
Wilson, A. J. C. (1942). *Determination of absolute from relative X-ray intensity data. Nature (London)*, **150**, 151–152.
Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature (London)*, **289**, 366–373.
Woolfson, M. M. (1956). *An improvement of the 'heavy-atom' method of solving crystal structures. Acta Cryst.* **9**, 804–810.
Woolfson, M. M. (1970). *An Introduction to X-ray Crystallography*. London: Cambridge University Press.
Wrinch, D. M. (1939). *The geometry of discrete vector maps. Philos. Mag.* **27**, 98–122.
Wunderlich, J. A. (1965). *A new expression for sharpening Patterson functions. Acta Cryst.* **19**, 200–202.
Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Away from the edge: SAD phasing from the sulfur anomalous signal measured in-house with chromium radiation*. *Acta Cryst.* D**59**, 1943–1957.
Yeates, T. O. & Rini, J. M. (1990). *Intensity-based domain refinement of oriented but unpositioned molecular replacement models*. *Acta Cryst*. A**46**, 352–359.
Zhang, K. Y. J. (1993). *SQUASH – combining constraints for macromolecular phase refinement and extension*. *Acta Cryst.* D**49**, 213–222.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *Phase improvement by iterative density modification*. In *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold, ch 15.1. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (2010). Vol. B, Chapter 2.4, pp. 282–296.

2.4. Isomorphous replacement and anomalous scattering

By M. Vijayan and S. Ramaseshan†

2.4.1. Introduction

Isomorphous replacement is among the earliest methods to be employed for crystal structure determination (Cork, 1927). The power of this method was amply demonstrated in the classical X-ray work of J. M. Robertson on phthalocyanine in the 1930s using centric data (Robertson, 1936; Robertson & Woodward, 1937). The structure determination of strychnine sulfate pentahydrate by Bijvoet and others provides an early example of the application of this method to acentric reflections (Bokhoven *et al.*, 1951). The usefulness of isomorphous replacement in the analysis of complex protein structures was demonstrated by Perutz and colleagues (Green *et al.*, 1954). This was closely followed by developments in the methodology for the application of isomorphous replacement to protein work (Harker, 1956; Blow & Crick, 1959) and rapidly led to the first ever structure solution of two related protein crystals, namely, those of myoglobin and haemoglobin (Kendrew *et al.*, 1960; Cullis *et al.*, 1961*b*). Since then isomorphous replacement has been the method of choice in macromolecular crystallography and most of the subsequent developments in and applications of this method have been concerned with biological macromolecules, mainly proteins (Blundell & Johnson, 1976; McPherson, 1982).

The application of anomalous-scattering effects has often developed in parallel with that of isomorphous replacement. Indeed, the two methods are complementary to a substantial extent and they are often treated together, as in this article. Although the most important effect of anomalous scattering, namely, the violation of Friedel's law, was experimentally observed as early as 1930 (Coster *et al.*, 1930), two decades elapsed before this effect was made use of for the first time by Bijvoet and his associates for the determination of the absolute configuration of asymmetric molecules as well as for phase evaluation (Bijvoet, 1949, 1954; Bijvoet *et al.*, 1951). Since then there has been a phenomenal spurt in the application of anomalous-scattering effects (Srinivasan, 1972; Ramaseshan & Abrahams, 1975; Vijayan, 1987). A quantitative formulation for the determination of phase angles using intensity differences between Friedel equivalents was derived by Ramachandran & Raman (1956), while Okaya & Pepinsky (1956) successfully developed a Patterson approach involving anomalous effects. The anomalous-scattering method of phase determination has since been used in the structure analysis of several structures, including those of a complex derivative of vitamin B_{12} (Dale *et al.*, 1963) and a small protein (Hendrickson & Teeter, 1981). In the meantime, the effect of changes in the real component of the dispersion correction as a function of the wavelength of the radiation used, first demonstrated by Mark & Szillard (1925), also received considerable attention. This effect, which is formally equivalent to that of isomorphous replacement, was demonstrated to be useful in structure determination (Ramaseshan *et al.*, 1957; Ramaseshan, 1963). Protein crystallographers have been quick to exploit anomalous-scattering effects (Rossmann, 1961; Kartha & Parthasarathy, 1965; North, 1965; Matthews, 1966; Hendrickson, 1979) and, as in the case of the isomorphous replacement method, the most useful applications of anomalous scattering during the last two decades have been perhaps in the field of macromolecular crystallography (Kartha, 1975; Watenpaugh *et al.*, 1975; Vijayan, 1981). In addition to anomalous scattering of X-rays, that of neutrons was also found to have interesting applications (Koetzle & Hamilton, 1975; Sikka & Rajagopal, 1975). More recently there has been a further revival in the development of anomalous-scattering methods with the advent of synchrotron radiation, particularly in view of the possibility of choosing any desired wavelength from a synchrotron-radiation source (Helliwell, 1984).

It is clear from the foregoing that the isomorphous replacement and the anomalous-scattering methods have a long and distinguished history. It is therefore impossible to do full justice to them in a comparatively short presentation like the present one. Several procedures for the application of these methods have been developed at different times. Many, although of considerable historical importance, are not extensively used at present for a variety of reasons. No attempt has been made to discuss them in detail here; the emphasis is primarily on the state of the art as it exists now. The available literature on isomorphous replacement and anomalous scattering is extensive. The reference list given at the end of this part is representative rather than exhaustive.

During the past few years, rapid developments have taken place in the isomorphous replacement and anomalous-scattering methods, particularly in the latter, as applied to macromolecular crystallography. These developments are described in detail in *International Tables for Crystallography*, Volume F (2001). Therefore, they have not been dealt with in this chapter. Significant developments in applications of direct methods to macromolecular crystallography have also occurred in recent years. A summary of these developments as well as the traditional direct methods on which the recent progress is based are presented in Chapter 2.2.

2.4.2. Isomorphous replacement method

2.4.2.1. Isomorphous replacement and isomorphous addition

Two crystals are said to be isomorphous if (*a*) both have the same space group and unit-cell dimensions and (*b*) the types and the positions of atoms in both are the same except for a replacement of one or more atoms in one structure with different types of atoms in the other (isomorphous replacement) or the presence of one or more additional atoms in one of them (isomorphous addition). Consider two crystal structures with identical space groups and unit-cell dimensions, one containing N atoms and the other M atoms. The N atoms in the first structure contain subsets P and Q whereas the M atoms in the second structure contain subsets P, Q' and R. The subset P is common to both structures in terms of atomic positions and atom types. The atomic positions are identical in subsets Q and Q', but at any given atomic position the atom type is different in Q and Q'. The subset R exists only in the second structure. If $\mathbf{F}_N$ and $\mathbf{F}_M$ denote the structure factors of the two structures for a given reflection,

$$\mathbf{F}_N = \mathbf{F}_P + \mathbf{F}_Q \qquad (2.4.2.1)$$

and

$$\mathbf{F}_M = \mathbf{F}_P + \mathbf{F}_{Q'} + \mathbf{F}_R, \qquad (2.4.2.2)$$

where the quantities on the right-hand side represent contributions from different subsets. From (2.4.2.1) and (2.4.2.2) we have

† Deceased.

Copyright © 2010 International Union of Crystallography

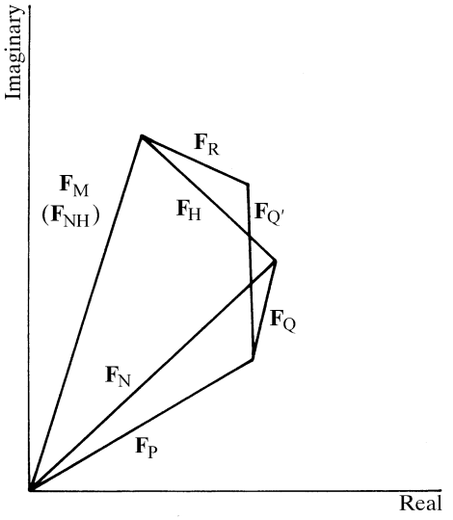


Fig. 2.4.2.1. Vector relationship between $\mathbf{F}_N$ and $\mathbf{F}_M$ ($\equiv \mathbf{F}_{NH}$).

$$\mathbf{F}_M - \mathbf{F}_N = \mathbf{F}_H = \mathbf{F}_{Q'} - \mathbf{F}_Q + \mathbf{F}_R. \qquad (2.4.2.3)$$

The above equations are illustrated in the Argand diagram shown in Fig. 2.4.2.1. $\mathbf{F}_Q$ and $\mathbf{F}_{Q'}$ would be collinear if all the atoms in Q were of the same type and those in Q' of another single type, as in the replacement of chlorine atoms in a structure by bromine atoms.

We have a case of 'isomorphous replacement' if $\mathbf{F}_R = 0$ ($\mathbf{F}_H = \mathbf{F}_{Q'} - \mathbf{F}_Q$) and a case of 'isomorphous addition' if $\mathbf{F}_Q = \mathbf{F}_{Q'} = 0$ ($\mathbf{F}_H = \mathbf{F}_R$). Once $\mathbf{F}_H$ is known, in addition to the magnitudes of $\mathbf{F}_N$ and $\mathbf{F}_M$, which can be obtained experimentally, the two cases can be treated in an equivalent manner in reciprocal space. In deference to common practice, the term 'isomorphous replacement' will be used to cover both cases. Also, in as much as $\mathbf{F}_M$ is the vector sum of $\mathbf{F}_N$ and $\mathbf{F}_H$, $\mathbf{F}_M$ and $\mathbf{F}_{NH}$ will be used synonymously. Thus

$$\mathbf{F}_M \equiv \mathbf{F}_{NH} = \mathbf{F}_N + \mathbf{F}_H. \qquad (2.4.2.4)$$

2.4.2.2. Single isomorphous replacement method

The number of replaceable (or 'added') atoms is usually small and they generally have high atomic numbers. Their positions are often determined by a Patterson synthesis of one type or another (see Chapter 2.3). It will therefore be assumed in the following discussion that $\mathbf{F}_H$ is known. Then it can be readily seen by referring to Fig. 2.4.2.2 that

$$\alpha_N = \alpha_H - \cos^{-1}\frac{F_{NH}^2 - F_N^2 - F_H^2}{2F_N F_H} = \alpha_H \pm \varphi; \qquad (2.4.2.5)$$

when φ is derived from its cosine function, it could obviously be positive or negative. Hence, there are two possible solutions for α_N. These two solutions are distributed symmetrically about $\mathbf{F}_H$. One of these would correspond to the correct value of α_N. Therefore, in general, the phase angle cannot be unambiguously determined using a pair of isomorphous crystals.

The twofold ambiguity in phase angle vanishes when the structures are centrosymmetric. $\mathbf{F}_{NH}$, $\mathbf{F}_N$ and $\mathbf{F}_H$ are all real in centric data and the corresponding phase angles are 0 or 180°. From (2.4.2.4)

$$F_{NH} \pm F_N = F_H. \qquad (2.4.2.6)$$

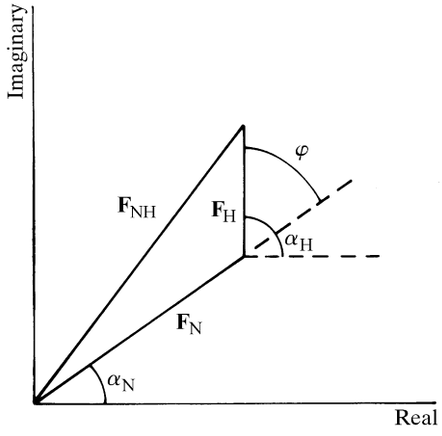


Fig. 2.4.2.2. Relationship between α_N, α_H and φ.

The sign of F_H is already known and the signs of F_{NH} and F_N can be readily determined from (2.4.2.6) (Robertson & Woodward, 1937).

When the data are acentric, the best one can do is to use both the possible phase angles simultaneously in a Fourier synthesis (Bokhoven *et al.*, 1951). This double-phased synthesis, which is equivalent to the isomorphous synthesis of Ramachandran & Raman (1959), contains the structure and its inverse when the replaceable atoms have a centrosymmetric distribution (Ramachandran & Srinivasan, 1970). When the distribution is noncentrosymmetric, however, the synthesis contains peaks corresponding to the structure and general background. Fourier syntheses computed using the single isomorphous replacement method of Blow & Rossmann (1961) and Kartha (1961) have the same properties. In this method, the phase angle is taken to be the average of the two possible solutions of α_N, which is always α_H or $\alpha_H + 180°$. Also, the Fourier coefficients are multiplied by cos φ, following arguments based on the Blow & Crick (1959) formulation of phase evaluation (see Section 2.4.4.4). Although Blow & Rossmann (1961) have shown that this method could yield interpretable protein Fourier maps, it is rarely used as such in protein crystallography as the Fourier maps computed using it usually have unacceptable background levels (Blundell & Johnson, 1976).

2.4.2.3. Multiple isomorphous replacement method

The ambiguity in α_N in a noncentrosymmetric crystal can be resolved only if at least two crystals isomorphous to it are available (Bokhoven *et al.*, 1951). We then have two equations of the type (2.4.2.5), namely,

$$\alpha_N = \alpha_{H1} \pm \varphi_1 \quad \text{and} \quad \alpha_N = \alpha_{H2} \pm \varphi_2, \qquad (2.4.2.7)$$

where subscripts 1 and 2 refer to isomorphous crystals 1 and 2, respectively. This is demonstrated graphically in Fig. 2.4.2.3 with the aid of the Harker (1956) construction. A circle is drawn with F_N as radius and the origin of the vector diagram as the centre. Two more circles are drawn with F_{NH1} and F_{NH2} as radii and the ends of vectors $-\mathbf{F}_{H1}$ and $-\mathbf{F}_{H2}$, respectively as centres. Each of these circles intersects the F_N circle at two points corresponding to the two possible solutions. One of the points of intersection is common and this point defines the correct value of α_N. With the assumption of perfect isomorphism and if errors are neglected, the phase circles corresponding to all the crystals would intersect at a common point if a number of isomorphous crystals were used for phase determination.

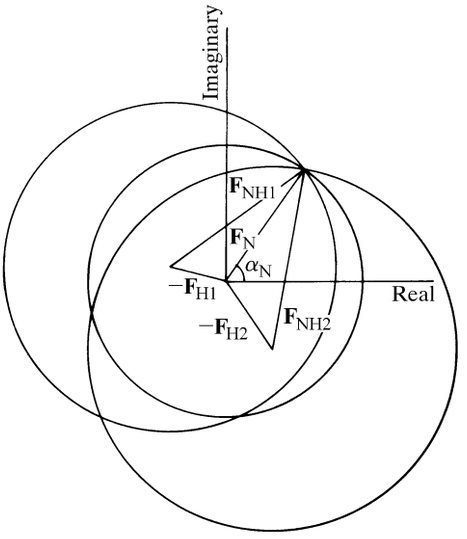


Fig. 2.4.2.3. Harker construction when two heavy-atom derivatives are available.

2.4.3. Anomalous-scattering method

2.4.3.1. Dispersion correction

Atomic scattering factors are normally calculated on the assumption that the binding energy of the electrons in an atom is negligible compared to the energy of the incident X-rays and the distribution of electrons is spherically symmetric. The transition frequencies within the atom are then negligibly small compared to the frequency of the radiation used and the scattering power of each electron in the atom is close to that of a free electron. When this assumption is valid, the atomic scattering factor is a real positive number and its value decreases as the scattering angle increases because of the finite size of the atom. When the binding energy of the electrons is appreciable, the atomic scattering factor at any given angle is given by

$$f_0 + f' + if'', \tag{2.4.3.1}$$

where f_0 is a real positive number and corresponds to the atomic scattering factor for a spherically symmetric collection of free electrons in the atom. The second and third terms are, respectively, referred to as the real and the imaginary components of the 'dispersion correction' (*IT* IV, 1974). f' is usually negative whereas f'' is positive. For any given atom, f'' is obviously 90° ahead of the real part of the scattering factor given by

$$f = f_0 + f'. \tag{2.4.3.2}$$

The variation of f' and f'' as a function of atomic number for two typical radiations is given in Fig. 2.4.3.1 (Srinivasan, 1972; Cromer, 1965). The dispersion effects are pronounced when an absorption edge of the atom concerned is in the neighbourhood of the wavelength of the incident radiation. Atoms with high atomic numbers have several absorption edges and the dispersion-correction terms in their scattering factors always have appreciable values. The values of f' and f'' do not vary appreciably with the angle of scattering as they are caused by core electrons confined to a very small volume around the nucleus. An atom is usually referred to as an anomalous scatterer if the dispersion-correction terms in its scattering factor have appreciable values. The effects on the structure factors or intensities of Bragg reflections resulting from dispersion corrections are referred to as anomalous-dispersion effects or anomalous-scattering effects.

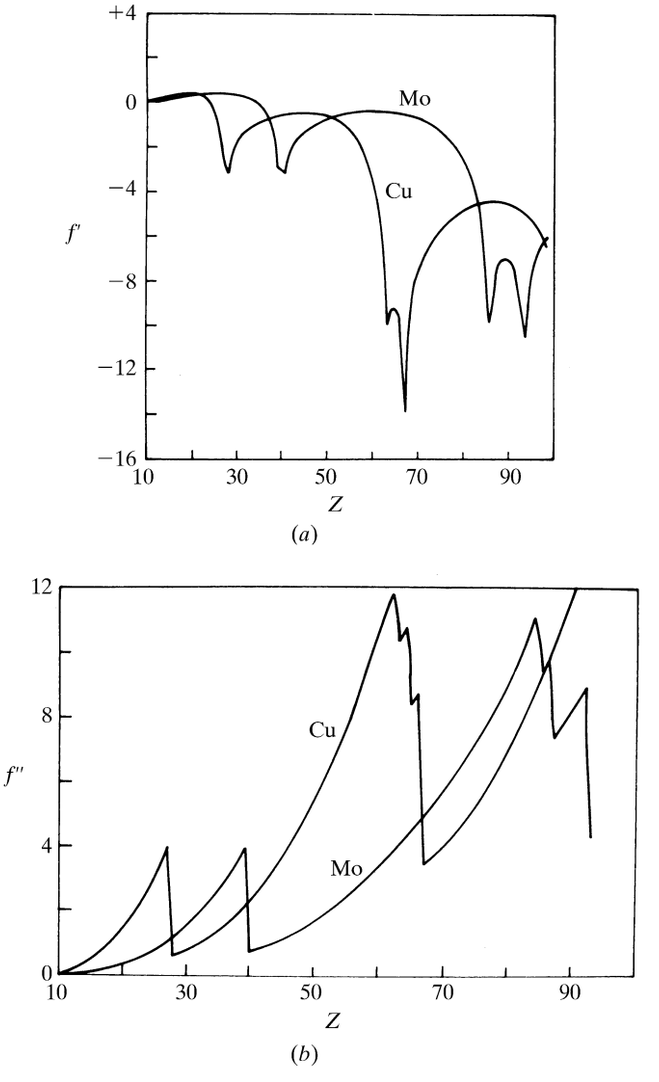


Fig. 2.4.3.1. Variation of (*a*) f' and (*b*) f'' as a function of atomic number for Cu $K\alpha$ and Mo $K\alpha$ radiations. Adapted from Fig. 3 of Srinivasan (1972).

2.4.3.2. Violation of Friedel's law

Consider a structure containing N atoms of which P are normal atoms and the remaining Q anomalous scatterers. Let $\mathbf{F}_P$ denote the contribution of the P atoms to the structure, and $\mathbf{F}_Q$ and $\mathbf{F}''_Q$ the real and imaginary components of the contribution of the Q atoms. The relation between the different contributions to a reflection $\mathbf{h}$ and its Friedel equivalent $-\mathbf{h}$ is illustrated in Fig. 2.4.3.2. For simplicity we assume here that all Q atoms are of the same type. The phase angle of $\mathbf{F}''_Q$ is then exactly 90° ahead of that of $\mathbf{F}_Q$. The structure factors of $\mathbf{h}$ and $-\mathbf{h}$ are denoted in the figure by $\mathbf{F}_N(+)$ and $\mathbf{F}_N(-)$, respectively. In the absence of anomalous scattering, or when the imaginary component of the dispersion correction is zero, the magnitudes of the two structure factors are equal and Friedel's law is obeyed; the phase angles have equal magnitudes, but opposite signs. As can be seen from Fig. 2.4.3.2, this is no longer true when $\mathbf{F}''_Q$ has a nonzero value. Friedel's law is then violated. A composite view of the vector relationship for $\mathbf{h}$ and $-\mathbf{h}$ can be obtained, as in Fig. 2.4.3.3, by reflecting the vectors corresponding to $-\mathbf{h}$ about the real axis of the vector diagram. $\mathbf{F}_P$ and $\mathbf{F}_Q$ corresponding to the two reflections superpose exactly, but $\mathbf{F}''_Q$ do not. $\mathbf{F}_N(+)$ and $\mathbf{F}_N(-)$ then have different magnitudes and phases.

It is easily seen that Friedel's law is obeyed in centric data even when anomalous scatterers are present. $\mathbf{F}_P$ and $\mathbf{F}_Q$ are then parallel to the real axis and $\mathbf{F}''_Q$ perpendicular to it. The vector sum of the three components is the same for $\mathbf{h}$ and $-\mathbf{h}$. It may, however, be noted that the phase angle of the structure factor is

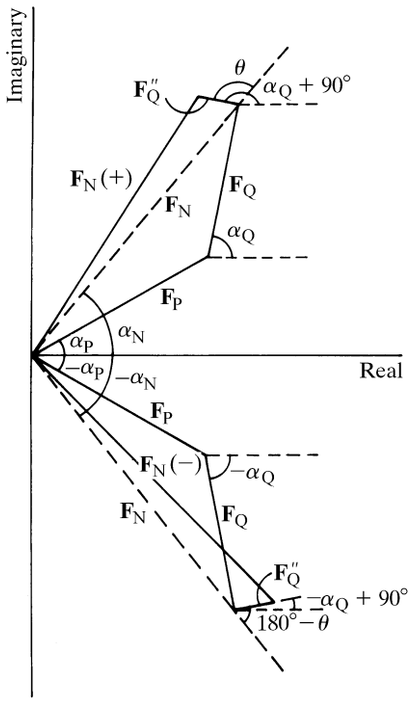


Fig. 2.4.3.2. Vector diagram illustrating the violation of Friedel's law when $\mathbf{F}''_Q \neq 0$.

then no longer 0 or 180°. Even when the structure is noncentrosymmetric, the effect of anomalous scattering in terms of intensity differences between Friedel equivalents varies from reflection to reflection. The difference between $\mathbf{F}_N(+)$ and $\mathbf{F}_N(-)$ is zero when $\alpha_P = \alpha_Q$ or $\alpha_Q + 180°$. The difference tends to the maximum possible value ($2F''_Q$) when $\alpha_P = \alpha_Q \pm 90°$.

Intensity differences between Friedel equivalents depend also on the ratio (in terms of number and scattering power) between anomalous and normal scatterers. Differences obviously do not occur when all the atoms are normal scatterers. On the other hand, a structure containing only anomalous scatterers of the same type also does not give rise to intensity differences. Expressions for intensity differences between Friedel equivalents have been derived by Zachariasen (1965) for the most general case of a structure containing normal as well as different types of anomalous scatterers. Statistical distributions of such differences under various conditions have also been derived (Parthasarathy & Srinivasan, 1964; Parthasarathy, 1967). It turns out that, with a single type of anomalous scatterer in the structure, the ratio

$$\frac{|F_N^2(+) - F_N^2(-)|}{F_N^2(+) + F_N^2(-)}$$

has a maximum mean value when the scattering powers of the anomalous scatterers and the normal scatterers are nearly the same (Srinivasan, 1972). Also, for a given ratio between the scattering powers, the smaller the number of anomalous scatterers, the higher is the mean ratio.

2.4.3.3. Friedel and Bijvoet pairs

The discussion so far has been concerned essentially with crystals belonging to space groups $P1$ and $P\bar{1}$. In the centrosymmetric space group, the crystal and the diffraction pattern have the same symmetry, namely, an inversion centre. In $P1$, however, the crystals are noncentrosymmetric while the diffraction pattern has an inversion centre, in the absence of anomalous scattering. When anomalous scatterers are present in the structure ($\mathbf{F}''_Q \neq 0$), Friedel's law breaks down and the diffraction pattern no longer has an inversion centre. Thus the diffraction

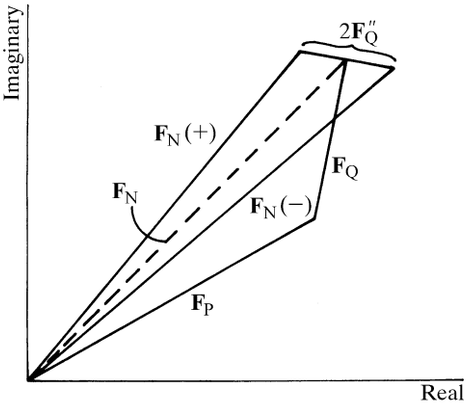


Fig. 2.4.3.3. A composite view of the vector relationship between $\mathbf{F}_N(+)$ and $\mathbf{F}_N(-)$.

pattern displays the same symmetry as that of the crystal in the presence of anomalous scattering. The same is true with higher-symmetry space groups also. For example, consider a crystal with space group $P222$, containing anomalous scatterers. The magnitudes of $\mathbf{F}_P$ are the same for all equivalent reflections; so are those of $\mathbf{F}_Q$ and $\mathbf{F}''_Q$. Their phase angles, however, differ from one equivalent to another, as can be seen from Table 2.4.3.1. When $\mathbf{F}''_Q = 0$, the magnitudes of the vector sum of $\mathbf{F}_P$ and $\mathbf{F}_Q$ are the same for all the equivalent reflections. The intensity pattern thus has point-group symmetry $2/m\ 2/m\ 2/m$. When $\mathbf{F}''_Q \neq 0$, the equivalent reflections can be grouped into two sets in terms of their intensities: hkl, $h\bar{k}\bar{l}$, $\bar{h}k\bar{l}$ and $\bar{h}\bar{k}l$; and $\bar{h}\bar{k}\bar{l}$, $\bar{h}kl$, $h\bar{k}l$ and $hk\bar{l}$. The equivalents belonging to the first group have the same intensity; so have the equivalents in the second group. But the two intensities are different. Thus the symmetry of the pattern is 222, the same as that of the crystal.

In general, under conditions of anomalous scattering, equivalent reflections generated by the symmetry elements in the crystal have intensities different from those of equivalent reflections generated by the introduction of an additional inversion centre in normal scattering. There have been suggestions that a reflection from the first group and another from the second group should be referred to as a 'Bijvoet pair' instead of a 'Friedel pair', when the two reflections are not inversely related. Most often, however, the terms are used synonymously. The same practice will be followed in this article.

2.4.3.4. Determination of absolute configuration

The determination of the absolute configuration of chiral molecules has been among the most important applications of anomalous scattering. Indeed, anomalous scattering is the only effective method for this purpose and the method, first used in the early 1950s (Peerdeman *et al.*, 1951), has been extensively employed in structural crystallography (Ramaseshan, 1963; Vos, 1975).

Many molecules, particularly biologically important ones, are chiral in that the molecular structure is not superimposable on its mirror image. Chirality (handedness) arises primarily on account of the presence of asymmetric carbon atoms in the molecule. A tetravalent carbon is asymmetric when the four atoms (or groups) bonded to it are all different from one another. The substituents can then have two distinct arrangements which are mirror images

Table 2.4.3.1. *Phase angles of different components of the structure factor in space group P222*

	Phase angle (°) of		
Reflection	$\mathbf{F}_P$	$\mathbf{F}_Q$	$\mathbf{F}''_Q$
hkl, $h\bar{k}\bar{l}$, $\bar{h}k\bar{l}$, $\bar{h}\bar{k}l$	α_P	α_Q	$90 + \alpha_Q$
$\bar{h}\bar{k}\bar{l}$, $\bar{h}kl$, $h\bar{k}l$, $hk\bar{l}$	$-\alpha_P$	$-\alpha_Q$	$90 - \alpha_Q$

of (or related by inversion to) each other. These optical isomers or enantiomers have the same chemical and physical properties except that they rotate the plane of polarization in opposite directions when polarized light passes through them. It is not, however, possible to calculate the sign of optical rotation, given the exact spatial arrangement or the 'absolute configuration' of the molecule. Therefore, one cannot distinguish between the possible enantiomorphic configurations of a given asymmetric molecule from measurements of optical rotation. This is also true of molecules with chiralities generated by overall asymmetric geometry instead of the presence of asymmetric carbon atoms in them.

Normal X-ray scattering does not distinguish between enantiomers. Two structures A (x_j, y_j, z_j) and B $(-x_j, -y_j, -z_j)$ $(j = 1, \ldots, N)$ obviously produce the same diffraction pattern on account of Friedel's law. The situation is, however, different when anomalous scatterers are present in the structure. The intensity difference between reflections $\mathbf{h}$ and $-\mathbf{h}$, or that between members of any Bijvoet pair, has the same magnitude, but opposite sign for structures A and B. If the atomic coordinates are known, the intensities of Bijvoet pairs can be readily calculated. The absolute configuration can then be determined, *i.e.* one can distinguish between A and B by comparing the calculated intensities with the observed ones for a few Bijvoet pairs with pronounced anomalous effects.

2.4.3.5. Determination of phase angles

An important application of anomalous scattering is in the determination of phase angles using Bijvoet differences (Ramachandran & Raman, 1956; Peerdeman & Bijvoet, 1956). From Figs. 2.4.3.2 and 2.4.3.3, we have

$$F_{\rm N}^2(+) = F_{\rm N}^2 + F_{\rm Q}''^2 + 2F_{\rm N}F_{\rm Q}'' \cos\theta \qquad (2.4.3.3)$$

and

$$F_{\rm N}^2(-) = F_{\rm N}^2 + F_{\rm Q}''^2 - 2F_{\rm N}F_{\rm Q}'' \cos\theta. \qquad (2.4.3.4)$$

Then

$$\cos\theta = \frac{F_{\rm N}^2(+) - F_{\rm N}^2(-)}{4F_{\rm N}F_{\rm Q}''}. \qquad (2.4.3.5)$$

In the above equations $F_{\rm N}$ may be approximated to $[F_{\rm N}(+) + F_{\rm N}(-)]/2$. Then θ can be evaluated from (2.4.3.5) except for the ambiguity in its sign. Therefore, from Fig. 2.4.3.2, we have

$$\alpha_{\rm N} = \alpha_{\rm Q} + 90^\circ \pm \theta. \qquad (2.4.3.6)$$

The phase angle thus has two possible values symmetrically distributed about $\mathbf{F}_{\rm Q}''$. Anomalous scatterers are usually heavy atoms and their positions can most often be determined by Patterson methods. $\alpha_{\rm Q}$ can then be calculated and the two possible values of $\alpha_{\rm N}$ for each reflection evaluated using (2.4.3.6).

In practice, the twofold ambiguity in phase angles can often be resolved in a relatively straightforward manner. As indicated earlier, anomalous scatterers usually have relatively high atomic numbers. The 'heavy-atom' phases calculated from their positions therefore contain useful information. For any given reflection, that phase angle which is closer to the heavy-atom phase, from the two phases calculated using (2.4.3.6), may be taken as the correct phase angle. This method has been successfully used in several structure determinations including that of a derivative of vitamin B_{12} (Dale *et al.*, 1963). The same method was also employed in a probabilistic fashion in the structure solution of a small protein (Hendrickson & Teeter, 1981). A method for obtaining a unique, but approximate, solution for phase angles from (2.4.3.6) has also been suggested (Srinivasan & Chacko, 1970). An accurate unique solution for phase angles can be obtained if one collects two sets of intensity data using two different wavelengths which have different dispersion-correction terms for the anomalous scatterers in the structure. Two equations of the type (2.4.3.6) are then available for each reflection and the solution common to both is obviously the correct phase angle. Different types of Patterson and Fourier syntheses can also be employed for structure solution using intensity differences between Bijvoet equivalents (Srinivasan, 1972; Okaya & Pepinsky, 1956; Pepinsky *et al.*, 1957).

An interesting situation occurs when the replaceable atoms in a pair of isomorphous structures are anomalous scatterers. The phase angles can then be uniquely determined by combining isomorphous replacement and anomalous-scattering methods. Such situations normally occur in protein crystallography and are discussed in Section 2.4.4.5.

2.4.3.6. Anomalous scattering without phase change

The phase determination, and hence the structure solution, outlined above relies on the imaginary component of the dispersion correction. Variation in the real component can also be used in structure analysis. In early applications of anomalous scattering, the real component of the dispersion correction was made use of to distinguish between atoms of nearly the same atomic numbers (Mark & Szillard, 1925; Bradley & Rodgers, 1934). For example, copper and manganese, with atomic numbers 29 and 25, respectively, are not easily distinguishable under normal X-ray scattering. However, the real components of the dispersion correction for the two elements are -1.129 and -3.367, respectively, when Fe $K\alpha$ radiation is used (*IT* IV, 1974). Therefore, the difference between the scattering factors of the two elements is accentuated when this radiation is used. The difference is more pronounced at high angles as the normal scattering factor falls off comparatively rapidly with increasing scattering angle whereas the dispersion-correction term does not.

The structure determination of $KMnO_4$ provides a typical example for the use of anomalous scattering without phase change in the determination of a centrosymmetric structure (Ramaseshan *et al.*, 1957; Ramaseshan & Venkatesan, 1957). f' and f'' for manganese for Cu $K\alpha$ radiation are -0.568 and 2.808, respectively. The corresponding values for Fe $K\alpha$ radiation are -3.367 and 0.481, respectively (*IT* IV, 1974). The data sets collected using the two radiations can now be treated as those arising from two perfectly isomorphous crystals. The intensity differences between a reflection in one set and the corresponding reflection in the other are obviously caused by the differences in the dispersion-correction terms. They can, however, be considered formally as intensity differences involving data from two perfectly isomorphous crystals. They can be used, as indeed they were, to determine the position of the manganese ion through an appropriate Patterson synthesis (see Section 2.4.4.2) and then to evaluate the signs of structure factors using (2.4.2.6) when the structure is centrosymmetric. When the structure is noncentrosymmetric, a twofold ambiguity exists in the phase angles in a manner analogous to that in the isomorphous replacement method. This ambiguity can be removed if the structure contains two different subsets of atoms $Q1$ and $Q2$ which, respectively, scatter radiations $\lambda_{\rm Q1}$ and $\lambda_{\rm Q2}$ anomalously. Data sets can then be collected with λ, which is scattered normally by all atoms, $\lambda_{\rm Q1}$ and $\lambda_{\rm Q2}$. The three sets can be formally treated as those from three perfectly isomorphous structures and the phase determination effected using (2.4.2.7) (Ramaseshan, 1963).

2.4.3.7. Treatment of anomalous scattering in structure refinement

The effect of anomalous scattering needs to be taken into account in the refinement of structures containing anomalous scatterers, if accurate atomic parameters are required. The effect of the real part of the dispersion correction is largely confined to the thermal parameters of anomalous scatterers. This effect can be eliminated by simply adding f' to the normal scattering factor of the anomalous scatterers.

The effects of the imaginary component of the dispersion correction are, however, more complex. These effects could lead to serious errors in positional parameters when the space group is polar, if data in the entire diffraction sphere are not used (Ueki *et al.*, 1966; Cruickshank & McDonald, 1967). For example, accessible data in a hemisphere are normally used for X-ray analysis when the space group is $P1$. If the hemisphere has say h positive, the x coordinates of all the atoms would be in error when the structure contains anomalous scatterers. The situation in other polar space groups has been discussed by Cruickshank & McDonald (1967). In general, in the presence of anomalous scattering, it is desirable to collect data for the complete sphere, if accurate structural parameters are required (Srinivasan, 1972).

Methods have been derived to correct for dispersion effects in observed data from centrosymmetric and noncentrosymmetric crystals (Patterson, 1963). The methods are empirical and depend upon the refined parameters at the stage at which corrections are applied. This is obviously an unsatisfactory situation and it has been suggested that the measured structure factors of Bijvoet equivalents should instead be treated as independent observations in structure refinement (Ibers & Hamilton, 1964). The effect of dispersion corrections needs to be taken into account to arrive at the correct scale and temperature factors also (Wilson, 1975; Gilli & Cruickshank, 1973).

2.4.4. Isomorphous replacement and anomalous scattering in protein crystallography

2.4.4.1. Protein heavy-atom derivatives

Perhaps the most spectacular applications of isomorphous replacement and anomalous-scattering methods have been in the structure solution of large biological macromolecules, primarily proteins. Since its first successful application on myoglobin and haemoglobin, the isomorphous replacement method, which is often used in conjunction with the anomalous-scattering method, has been employed in the solution of scores of proteins. The application of this method involves the preparation of protein heavy-atom derivatives, *i.e.* the attachment of heavy atoms like mercury, uranium and lead, or chemical groups containing them, to protein crystals in a coherent manner without changing the conformation of the molecules and their crystal packing. This is only rarely possible in ordinary crystals as the molecules in them are closely packed. Protein crystals, however, contain large solvent regions and isomorphous derivatives can be prepared by replacing the disordered solvent molecules by heavy-atom-containing groups without disturbing the original arrangement of protein molecules.

2.4.4.2. Determination of heavy-atom parameters

For any given reflection, the structure factor of the native protein crystal ($\mathbf{F}_{\rm N}$), that of a heavy-atom derivative ($\mathbf{F}_{\rm NH}$), and the contribution of the heavy atoms in that derivative ($\mathbf{F}_{\rm H}$) are related by the equation

$$\mathbf{F}_{\rm NH} = \mathbf{F}_{\rm N} + \mathbf{F}_{\rm H}. \qquad (2.4.4.1)$$

The value of $\mathbf{F}_{\rm H}$ depends not only on the positional and thermal parameters of the heavy atoms, but also on their occupancy factors, because, at a given position, the heavy atom may not often be present in all the unit cells. For example, if the heavy atom is present at a given position in only half the unit cells in the crystal, then the occupancy factor of the site is said to be 0.5.

For the successful determination of the heavy-atom parameters, as also for the subsequent phase determination, the data sets from the native and the derivative crystals should have the same relative scale. The different data sets should also have the same overall temperature factor. Different scaling procedures have been suggested (Blundell & Johnson, 1976) and, among them, the following procedure, based on Wilson's (1942) statistics, appears to be the most feasible in the early stages of structure analysis.

Assuming that the data from the native and the derivative crystals obey Wilson's statistics, we have, for any range of $\sin^2\theta/\lambda^2$,

$$\ln\left\{\frac{\sum f_{{\rm N}j}^2}{\langle F_{\rm N}^2\rangle}\right\} = \ln K_{\rm N} + 2B_{\rm N}\frac{\sin^2\theta}{\lambda^2} \qquad (2.4.4.2)$$

and

$$\ln\left\{\frac{\sum f_{{\rm N}j}^2 + \sum f_{{\rm H}j}^2}{\langle F_{\rm NH}^2\rangle}\right\} = \ln K_{\rm NH} + 2B_{\rm NH}\frac{\sin^2\theta}{\lambda^2}, \qquad (2.4.4.3)$$

where $f_{{\rm N}j}$ and $f_{{\rm H}j}$ refer to the atomic scattering factors of protein atoms and heavy atoms, respectively. $K_{\rm N}$ and $K_{\rm NH}$ are the scale factors to be applied to the intensities from the native and the derivative crystals, respectively, and $B_{\rm N}$ and $B_{\rm NH}$ the temperature factors of the respective structure factors. Normally one would be able to derive the absolute scale factor and the temperature factor for both the data sets from (2.4.4.2) and (2.4.4.3) using the well known Wilson plot. The data from protein crystals, however, do not follow Wilson's statistics as protein molecules contain highly non-random features. Therefore, in practice, it is difficult to fit a straight line through the points in a Wilson plot, thus rendering the parameters derived from it unreliable. (2.4.4.2) and (2.4.4.3) can, however, be used in a different way. From the two equations we obtain

$$\ln\left\{\frac{\sum f_{{\rm N}j}^2 + \sum f_{{\rm H}j}^2}{\sum f_{{\rm N}j}^2}\cdot\frac{\langle F_{\rm N}^2\rangle}{\langle F_{\rm NH}^2\rangle}\right\} = \ln\left(\frac{K_{\rm NH}}{K_{\rm N}}\right) + 2(B_{\rm NH} - B_{\rm N})\frac{\sin^2\theta}{\lambda^2}. \qquad (2.4.4.4)$$

The effects of structural non-randomness in the crystals obviously cancel out in (2.4.4.4). When the left-hand side of (2.4.4.4) is plotted against $(\sin^2\theta)/\lambda^2$, it is called a comparison or difference Wilson plot. Such plots yield the ratio between the scales of the derivative and the native data, and the additional temperature factor of the derivative data. Initially, the number and the occupancy factors of heavy-atom sites are unknown, and are roughly estimated from intensity differences to evaluate $\sum f_{{\rm H}j}^2$. These estimates usually undergo considerable revision in the course of the determination and the refinement of heavy-atom parameters.

At first, heavy-atom positions are most often determined by Patterson syntheses of one type or another. Such syntheses are discussed in some detail elsewhere in Chapter 2.3. They are therefore discussed here only briefly.

Equation (2.4.2.6) holds when the data are centric. $F_{\rm H}$ is usually small compared to $F_{\rm N}$ and $F_{\rm NH}$, and the minus sign is then relevant on the left-hand side of (2.4.2.6). Thus the difference between the magnitudes of $\mathbf{F}_{\rm NH}$ and $\mathbf{F}_{\rm N}$, which can be obtained

experimentally, normally gives a correct estimate of the magnitude of $\mathbf{F}_H$ for most reflections. Then a Patterson synthesis with $(F_{NH} - F_N)^2$ as coefficients corresponds to the distribution of vectors between heavy atoms, when the data are centric. But proteins are made up of L-amino acids and hence cannot crystallize in centrosymmetric space groups. However, many proteins crystallize in space groups with centrosymmetric projections. The centric data corresponding to these projections can then be used for determining heavy-atom positions through a Patterson synthesis of the type outlined above.

The situation is more complex for three-dimensional acentric data. It has been shown (Rossmann, 1961) that

$$(F_{NH} - F_N)^2 \simeq F_H^2 \cos^2(\alpha_{NH} - \alpha_H) \qquad (2.4.4.5)$$

when F_H is small compared to F_{NH} and F_N. Patterson synthesis with $(F_{NH} - F_N)^2$ as coefficients would, therefore, give an approximation to the heavy-atom vector distribution. An isomorphous difference Patterson synthesis of this type has been used extensively in protein crystallography to determine heavy-atom positions. The properties of this synthesis have been extensively studied (Ramachandran & Srinivasan, 1970; Rossmann, 1960; Phillips, 1966; Dodson & Vijayan, 1971) and it has been shown that this Patterson synthesis would provide a good approximation to the heavy-atom vector distribution even when F_H is large compared to F_N (Dodson & Vijayan, 1971).

As indicated earlier (see Section 2.4.3.1), heavy atoms are always anomalous scatterers, and the structure factors of any given reflection and its Friedel equivalent from a heavy-atom derivative have unequal magnitudes. If these structure factors are denoted by $\mathbf{F}_{NH}(+)$ and $\mathbf{F}_{NH}(-)$ and the real component of the heavy-atom contributions (including the real component of the dispersion correction) by $\mathbf{F}_H$, then it can be shown (Kartha & Parthasarathy, 1965) that

$$\left(\frac{k}{2}\right)^2 [F_{NH}(+) - F_{NH}(-)]^2 = F_H^2 \sin^2(\alpha_{NH} - \alpha_H), \qquad (2.4.4.6)$$

where $k = (f_H + f'_H)/f''_H$. Here it has been assumed that all the anomalous scatterers are of the same type with atomic scattering factor f_H and dispersion-correction terms f'_H and f''_H. A Patterson synthesis with the left-hand side of (2.4.4.6) as coefficients would also yield the vector distribution corresponding to the heavy-atom positions (Rossmann, 1961; Kartha & Parthasarathy, 1965). However, $F_{NH}(+) - F_{NH}(-)$ is a small difference between two large quantities and is liable to be in considerable error. Patterson syntheses of this type are therefore rarely used to determine heavy-atom positions.

It is interesting to note (Kartha & Parthasarathy, 1965) that addition of (2.4.4.5) and (2.4.4.6) readily leads to

$$(F_{NH} - F_N)^2 + \left(\frac{k}{2}\right)^2 [F_{NH}(+) - F_{NH}(-)]^2 \simeq F_H^2. \qquad (2.4.4.7)$$

Thus, the magnitude of the heavy-atom contribution can be estimated if intensities of Friedel equivalents have been measured from the derivative crystal. F_{NH} is then not readily available, but to a good approximation

$$F_{NH} = [F_{NH}(+) + F_{NH}(-)]/2. \qquad (2.4.4.8)$$

A different and more accurate expression for estimating F_H^2 from isomorphous and anomalous differences was derived by Matthews (1966). According to a still more accurate expression derived by Singh & Ramaseshan (1966),

$$\begin{aligned} F_H^2 &= F_{NH}^2 + F_N^2 - 2F_{NH}F_N \cos(\alpha_N - \alpha_{NH}) \\ &= F_{NH}^2 + F_N^2 \pm 2F_{NH}F_N \\ &\quad \times (1 - \{k[F_{NH}(+) - F_{NH}(-)]/2F_N\}^2)^{1/2}. \end{aligned} \qquad (2.4.4.9)$$

The lower estimate in (2.4.4.9) is relevant when $|\alpha_N - \alpha_{NH}| < 90^\circ$ and the upper estimate is relevant when $|\alpha_N - \alpha_{NH}| > 90^\circ$. The lower and the upper estimates may be referred to as F_{HLE} and F_{HUE}, respectively. It can be readily shown (Dodson & Vijayan, 1971) that the lower estimate would represent the correct value of F_H for a vast majority of reflections. Thus, a Patterson synthesis with F_{HLE}^2 as coefficients would yield the vector distribution of heavy atoms in the derivative. Such a synthesis would normally be superior to those with the left-hand sides of (2.4.4.5) and (2.4.4.6) as coefficients. However, when the level of heavy-atom substitution is low, the anomalous differences are also low and susceptible to large percentage errors. In such a situation, a synthesis with $(F_{NH} - F_N)^2$ as coefficients is likely to yield better results than that with F_{HLE}^2 as coefficients (Vijayan, 1981).

Direct methods employing different methodologies have also been used successfully for the determination of heavy-atom positions (Navia & Sigler, 1974). These methods, developed primarily for the analysis of smaller structures, have not yet been successful in *a priori* analysis of protein structures. The very size of protein structures makes the probability relations used in these methods weak. In addition, data from protein crystals do not normally extend to high enough angles to permit resolution of individual atoms in the structure and the feasibility of using many of the currently popular direct-method procedures in such a situation has been a topic of much discussion. The heavy atoms in protein derivative crystals, however, are small in number and are normally situated far apart from one another. They are thus expected to be resolved even when low-resolution X-ray data are used. In most applications, the magnitudes of the differences between F_{NH} and F_N are formally considered as the 'observed structure factors' of the heavy-atom distribution and conventional direct-method procedures are then applied to them.

Once the heavy-atom parameters in one or more derivatives have been determined, approximate protein phase angles, α_N's, can be derived using methods described later. These phase angles can then be readily used to determine the heavy-atom parameters in a new derivative employing a difference Fourier synthesis with coefficients

$$(F_{NH} - F_N) \exp(i\alpha_N). \qquad (2.4.4.10)$$

Such syntheses are also used to confirm and to improve upon the information on heavy-atom parameters obtained through Patterson or direct methods. They are obviously very powerful when centric data corresponding to centrosymmetric projections are used. The synthesis yields satisfactory results even when the data are acentric although the difference Fourier technique becomes progressively less powerful as the level of heavy-atom substitution increases (Dodson & Vijayan, 1971).

While the positional parameters of heavy atoms can be determined with a reasonable degree of confidence using the above-mentioned methods, the corresponding temperature and occupancy factors cannot. Rough estimates of the latter are usually made from the strength and the size of appropriate peaks in difference syntheses. The estimated values are then refined, along with the positional parameters, using the techniques outlined below.

2.4.4.3. Refinement of heavy-atom parameters

The least-squares method with different types of minimization functions is used for refining the heavy-atom parameters, including the occupancy factors. The most widely used method (Dickerson *et al.*, 1961; Muirhead *et al.*, 1967; Dickerson *et al.*, 1968) involves the minimization of the function

$$\varphi = \sum w(F_{\rm NH} - |\mathbf{F}_{\rm N} + \mathbf{F}_{\rm H}|)^2, \qquad (2.4.4.11)$$

where the summation is over all the reflections and w is the weight factor associated with each reflection. Here $F_{\rm NH}$ is the observed magnitude of the structure factor for the particular derivative and $\mathbf{F}_{\rm N} + \mathbf{F}_{\rm H}$ is the calculated structure factor. The latter obviously depends upon the protein phase angle $\alpha_{\rm N}$, and the magnitude and the phase angle of $\mathbf{F}_{\rm H}$ which are in turn dependent on the heavy-atom parameters. Let us assume that we have three derivatives A, B and C, and that we have already determined the heavy-atom parameters HA_i, HB_i and HC_i. Then,

$$\begin{aligned} \mathbf{F}_{\rm HA} &= \mathbf{F}_{\rm HA}(HA_i) \\ \mathbf{F}_{\rm HB} &= \mathbf{F}_{\rm HB}(HB_i) \\ \mathbf{F}_{\rm HC} &= \mathbf{F}_{\rm HC}(HC_i). \end{aligned} \qquad (2.4.4.12)$$

A set of approximate protein phase angles is first calculated, employing methods described later, making use of the unrefined heavy-atom parameters. These phase angles are used to construct $\mathbf{F}_{\rm N} + \mathbf{F}_{\rm H}$ for each derivative. (2.4.4.11) is then minimized, separately for each derivative, by varying HA_i for derivative A, HB_i for derivative B, and HC_i for derivative C. The refined values of HA_i, HB_i and HC_i are subsequently used to calculate a new set of protein phase angles. Alternate cycles of parameter refinement and phase-angle calculation are carried out until convergence is reached. The progress of refinement may be monitored by computing an R factor defined as (Kraut *et al.*, 1962)

$$R_{\rm K} = \frac{\sum |F_{\rm NH} - |\mathbf{F}_{\rm N} + \mathbf{F}_{\rm H}||}{F_{\rm NH}}. \qquad (2.4.4.13)$$

The above method has been successfully used for the refinement of heavy-atom parameters in the X-ray analysis of many proteins. However, it has one major drawback in that the refined parameters in one derivative are dependent on those in other derivatives through the calculation of protein phase angles. Therefore, it is important to ensure that the derivative, the heavy-atom parameters of which are being refined, is omitted from the phase-angle calculation (Blow & Matthews, 1973). Even when this is done, serious problems might arise when different derivatives are related by common sites. In practice, the occupancy factors of the common sites tend to be overestimated compared to those of the others (Vijayan, 1981; Dodson & Vijayan, 1971). Yet another factor which affects the occupancy factors is the accuracy of the phase angles. The inclusion of poorly phased reflections tends to result in the underestimation of occupancy factors. It is therefore advisable to omit from refinement cycles reflections with figures of merit less than a minimum threshold value or to assign a weight proportional to the figure of merit (as defined later) to each term in the minimization function (Dodson & Vijayan, 1971; Blow & Matthews, 1973).

If anomalous-scattering data from derivative crystals are available, the values of $F_{\rm H}$ can be estimated using (2.4.4.7) or (2.4.4.9) and these can be used as the 'observed' magnitudes of the heavy-atom contributions for the refinement of heavy-atom parameters, as has been done by many workers (Watenpaugh *et al.*, 1975; Vijayan, 1981; Kartha, 1965). If (2.4.4.9) is used for estimating $F_{\rm H}$, the minimization function has the form

$$\varphi = \sum w(F_{\rm HLE} - F_{\rm H})^2. \qquad (2.4.4.14)$$

The progress of refinement may be monitored using a reliability index defined as

$$R = \frac{\sum |F_{\rm HLE} - F_{\rm H}|}{\sum F_{\rm HLE}}. \qquad (2.4.4.15)$$

The major advantage of using $F_{\rm HLE}$'s in refinement is that the heavy-atom parameters in each derivative can now be refined independently of all other derivatives. Care should, however, be taken to omit from calculations all reflections for which $F_{\rm HUE}$ is likely to be the correct estimate of $F_{\rm H}$. This can be achieved in practice by excluding from least-squares calculations all reflections for which $F_{\rm HUE}$ has a value less than the maximum expected value of $F_{\rm H}$ for the given derivative (Vijayan, 1981; Dodson & Vijayan, 1971).

A major problem associated with this refinement method is concerned with the effect of experimental errors on refined parameters. The values of $F_{\rm NH}(+) - F_{\rm NH}(-)$ are often comparable to the experimental errors associated with $F_{\rm NH}(+)$ and $F_{\rm NH}(-)$. In such a situation, even random errors in $F_{\rm NH}(+)$ and $F_{\rm NH}(-)$ tend to increase systematically the observed difference between them (Dodson & Vijayan, 1971). In (2.4.4.7) and (2.4.4.9), this difference is multiplied by k or $k/2$, a quantity much greater than unity, and then squared. This could lead to the systematic overestimation of $F_{\rm HLE}$'s and the consequent overestimation of occupancy factors. The situation can be improved by employing empirical values of k, evaluated using the relation (Kartha & Parthasarathy, 1965; Matthews, 1966)

$$k = \frac{2\sum |F_{\rm NH} - F_{\rm N}|}{\sum |F_{\rm NH}(+) - F_{\rm NH}(-)|}, \qquad (2.4.4.16)$$

for estimating $F_{\rm HLE}$ or by judiciously choosing the weighting factors in (2.4.4.14) (Dodson & Vijayan, 1971). The use of a modified form of $F_{\rm HLE}$, arrived at through statistical considerations, along with appropriate weighting factors, has also been advocated (Dodson *et al.*, 1975).

When the data are centric, (2.4.4.9) reduces to

$$F_{\rm H} = F_{\rm NH} \pm F_{\rm N}. \qquad (2.4.4.17)$$

Here, again, the lower estimate most often corresponds to the correct value of $F_{\rm H}$. (2.4.4.17) does not involve $F_{\rm NH}(+) - F_{\rm NH}(-)$ which, as indicated earlier, is prone to substantial error. Therefore, $F_{\rm H}$'s estimated using centric data are more reliable than those estimated using acentric data. Consequently, centric reflections, when available, are extensively used for the refinement of heavy-atom parameters. It may also be noted that in conditions under which $F_{\rm HLE}$ corresponds to the correct estimate of $F_{\rm H}$, minimization functions (2.4.4.11) and (2.4.4.14) are identical for centric data.

A Patterson function correlation method with a minimization function of the type

$$\varphi = \sum w[(F_{\rm NH} - F_{\rm N})^2 - F_{\rm H}^2]^2 \qquad (2.4.4.18)$$

was among the earliest procedures suggested for heavy-atom-parameter refinement (Rossmann, 1960). This procedure would obviously work well when centric reflections are used. A modified version of this procedure, in which the origins of the Patterson functions are removed from the correlation, and

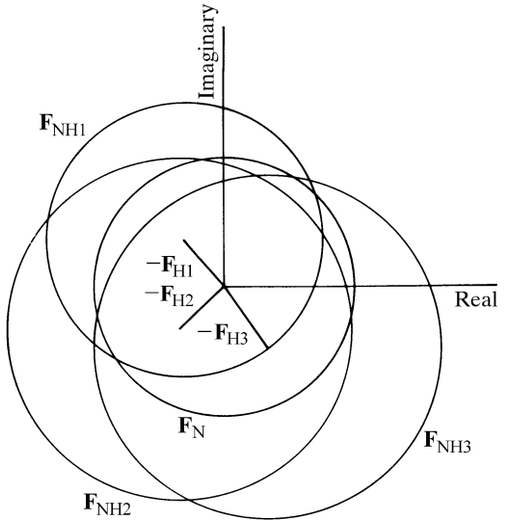


Fig. 2.4.4.1. Distribution of intersections in the Harker construction under non-ideal conditions.

centric and acentric data are treated separately, has been proposed (Terwilliger & Eisenberg, 1983).

2.4.4.4. Treatment of errors in phase evaluation: Blow and Crick formulation

As shown in Section 2.4.2.3, ideally, protein phase angles can be evaluated if two isomorphous heavy-atom derivatives are available. However, in practice, conditions are far from ideal on account of several factors such as imperfect isomorphism, errors in the estimation of heavy-atom parameters, and the experimental errors in the measurement of intensity from the native and the derivative crystals. It is therefore desirable to use as many derivatives as are available for phase determination. When isomorphism is imperfect and errors exist in data and heavy-atom parameters, all the circles in a Harker diagram would not intersect at a single point; instead, there would be a distribution of intersections, such as that illustrated in Fig. 2.4.4.1. Consequently, a unique solution for the phase angle cannot be deduced.

The statistical procedure for computing protein phase angles using multiple isomorphous replacement (MIR) was derived by Blow & Crick (1959). In their treatment, Blow and Crick assume, for mathematical convenience, that all errors, including those arising from imperfect isomorphism, could be considered as residing in the magnitudes of the derivative structure factors only. They further assume that these errors could be described by a Gaussian distribution. With these simplifying assumptions, the statistical procedure for phase determination could be derived in the following manner.

Consider the vector diagram, shown in Fig. 2.4.4.2, for a reflection from the ith derivative for an arbitrary value α for the protein phase angle. Then,

$$D_{Hi}(\alpha) = [F_N^2 + F_{Hi}^2 + 2F_N F_{Hi}\cos(\alpha_{Hi} - \alpha)]^{1/2}. \qquad (2.4.4.19)$$

If α corresponds to the true protein phase angle α_N, then D_{Hi} coincides with F_{NHi}. The amount by which $D_{Hi}(\alpha)$ differs from F_{NHi}, namely,

$$\xi_{Hi}(\alpha) = F_{NHi} - D_{Hi}(\alpha), \qquad (2.4.4.20)$$

is a measure of the departure of α from α_N. ξ is called the lack of closure. The probability for α being the correct protein phase angle could now be defined as

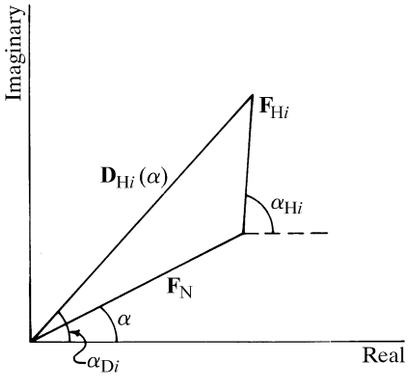


Fig. 2.4.4.2. Vector diagram indicating the calculated structure factor, $\mathbf{D}_{Hi}(\alpha)$, of the ith heavy-atom derivative for an arbitrary value α for the phase angle of the structure factor of the native protein.

$$P_i(\alpha) = N_i \exp[-\xi_{Hi}^2(\alpha)/2E_i^2], \qquad (2.4.4.21)$$

where N_i is the normalization constant and E_i is the estimated r.m.s. error. The methods for estimating E_i will be outlined later. When several derivatives are used for phase determination, the total probability of the phase angle α being the protein phase angle would be

$$P(\alpha) = \prod P_i(\alpha) = N \exp\left\{-\sum_i [\xi_{Hi}^2(\alpha)/2E_i^2]\right\}, \qquad (2.4.4.22)$$

where the summation is over all the derivatives. A typical distribution of $P(\alpha)$ plotted around a circle of unit radius is shown in Fig. 2.4.4.3. The phase angle corresponding to the highest value of $P(\alpha)$ would obviously be the most probable protein phase, α_M, of the given reflection. The most probable electron-density distribution is obtained if each F_N is associated with the corresponding α_M in a Fourier synthesis.

Blow and Crick suggested a different way of using the probability distribution. In Fig. 2.4.4.3, the centroid of the probability distribution is denoted by P. The polar coordinates of P are m and α_B, where m, a fractional positive number with a maximum value of unity, and α_B are referred to as the 'figure of merit' and the 'best phase', respectively. One can then compute a 'best Fourier' with coefficients

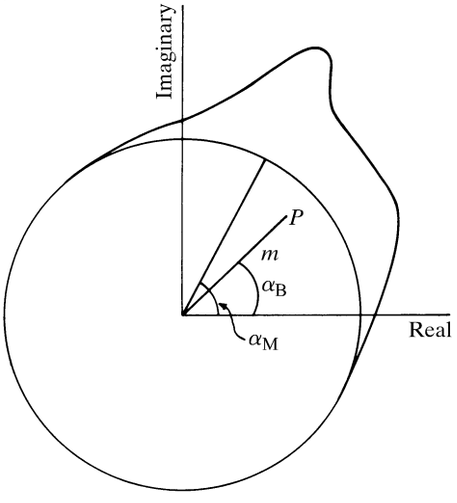


Fig. 2.4.4.3. The probability distribution of the protein phase angle. The point P is the centroid of the distribution.

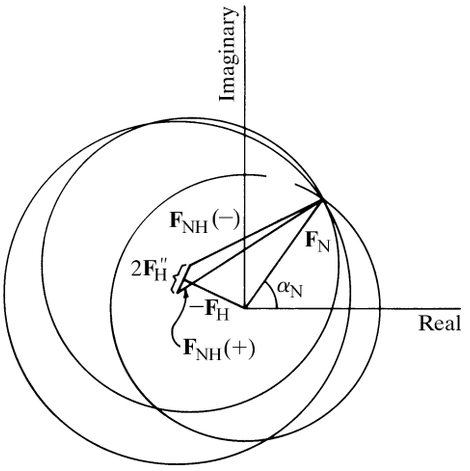


Fig. 2.4.4.4. Harker construction using anomalous-scattering data from a single derivative.

$$mF_N \exp(i\alpha_B).$$

The best Fourier is expected to provide an electron-density distribution with the lowest r.m.s. error. The figure of merit and the best phase are usually calculated using the equations

$$\begin{aligned} m\cos\alpha_B &= \sum_i P(\alpha_i)\cos(\alpha_i)/\sum_i P(\alpha_i) \\ m\sin\alpha_B &= \sum_i P(\alpha_i)\sin(\alpha_i)/\sum_i P(\alpha_i), \end{aligned} \qquad (2.4.4.23)$$

where $P(\alpha_i)$ are calculated, say, at 5° intervals (Dickerson *et al.*, 1961). The figure of merit is statistically interpreted as the cosine of the expected error in the calculated phase angle and it is obviously a measure of the precision of phase determination. In general, m is high when α_M and α_B are close to each other and low when they are far apart.

2.4.4.5. Use of anomalous scattering in phase evaluation

When anomalous-scattering data have been collected from derivative crystals, $F_{NH}(+)$ and $F_{NH}(-)$ can be formally treated as arising from two independent derivatives. The corresponding Harker diagram is shown in Fig. 2.4.4.4. Thus, in principle, protein phase angles can be determined using a single derivative when anomalous-scattering effects are also made use of. It is interesting to note that the information obtained from isomorphous differences, $F_{NH} - F_N$, and that obtained from anomalous differences, $F_{NH}(+) - F_{NH}(-)$, are complementary. The isomorphous difference for any given reflection is a maximum when F_N and F_H are parallel or antiparallel. The anomalous difference is then zero, if all the anomalous scatterers are of the same type, and α_N is determined uniquely on the basis of the isomorphous difference. The isomorphous difference decreases and the anomalous difference increases as the inclination between $\mathbf{F}_N$ and $\mathbf{F}_H$ increases. The isomorphous difference tends to be small and the anomalous difference tends to have the maximum possible value when $\mathbf{F}_N$ and $\mathbf{F}_H$ are perpendicular to each other. The anomalous difference then has the predominant influence in determining the phase angle.

Although isomorphous and anomalous differences have a complementary role in phase determination, their magnitudes are obviously unequal. Therefore, when $F_{NH}(+)$ and $F_{NH}(-)$ are treated as arising from two derivatives, the effect of anomalous differences on phase determination would be only marginal as, for any given reflection, $F_{NH}(+) - F_{NH}(-)$ is usually much smaller than $F_{NH} - F_N$. However, the magnitude of the error in the anomalous difference would normally be much smaller than that in the corresponding isomorphous difference. Firstly, the former is obviously free from the effects of imperfect isomorphism. Secondly, $F_{NH}(+)$ and $F_{NH}(-)$ are expected to have the same systematic errors as they are measured from the same crystal. These errors are eliminated in the difference between the two quantities. Therefore, as pointed out by North (1965), the r.m.s. error used for anomalous differences should be much smaller than that used for isomorphous differences. Denoting the r.m.s. error in anomalous differences by E', the new expression for the probability distribution of protein phase angle may be written as

$$\begin{aligned} P_i(\alpha) = N_i \exp[-\xi^2_{Hi}(\alpha)/2E_i^2] \\ \times \exp\{-[\Delta H_i - \Delta H_{ical}(\alpha)]^2/2E_i'^2\}, \end{aligned} \qquad (2.4.4.24)$$

where

$$\Delta H_i = F_{NHi}(+) - F_{NHi}(-)$$

and

$$\Delta H_{ical}(\alpha) = 2F''_{Hi}\sin(\alpha_{Di} - \alpha_{Hi}).$$

Here α_{Di} is the phase angle of $D_{Hi}(\alpha)$ [see (2.4.4.19) and Fig. 2.4.4.2]. $\Delta H_{ical}(\alpha)$ is the anomalous difference calculated for the assumed protein phase angle α. F_{NHi} may be taken as the average of $F_{NHi}(+)$ and $F_{NHi}(-)$ for calculating $\xi^2_{Hi}(\alpha)$ using (2.4.4.20).

2.4.4.6. Estimation of r.m.s. error

Perhaps the most important parameters that control the reliability of phase evaluation using the Blow and Crick formulation are the isomorphous r.m.s. error E_i and the anomalous r.m.s. error E'_i. For a given derivative, the sharpness of the peak in the phase probability distribution obviously depends upon the value of E and that of E' when anomalous-scattering data have also been used. When several derivatives are used, an overall underestimation of r.m.s. errors leads to artifically sharper peaks, the movement of α_B towards α_M, and deceptively high figures of merit. Opposite effects result when E's are overestimated. Underestimation or overestimation of the r.m.s. error in the data from a particular derivative leads to distortions in the relative contribution of that derivative to the overall phase probability distributions. It is therefore important that the r.m.s. error in each derivative is correctly estimated.

Centric reflections, when present, obviously provide the best means for evaluating E using the expression

$$E^2 = \sum_n (|F_{NH} \pm F_N| - F_N)^2/n. \qquad (2.4.4.25)$$

As suggested by Blow & Crick (1959), values of E thus estimated can be used for acentric reflections as well. Once a set of approximate protein phase angles is available, E_i can be calculated as the r.m.s. lack of closure corresponding to α_B [*i.e.* $\alpha = \alpha_B$ in (2.4.4.20)] (Kartha, 1976). E'_i can be similarly evaluated as the r.m.s. difference between the observed anomalous difference and the anomalous difference calculated for α_B [see (2.4.4.24)]. Normally, the value of E'_i is about a third of that of E_i (North, 1965).

A different method, outlined below, can also be used to evaluate E and E' when anomalous scattering is present (Vijayan, 1981; Adams, 1968). From Fig. 2.4.2.2, we have

$$\cos\psi = (F_{\rm NH}^2 + F_{\rm H}^2 - F_{\rm N}^2)/2F_{\rm NH}F_{\rm H} \qquad (2.4.4.26)$$

and

$$F_{\rm N}^2 = F_{\rm NH}^2 + F_{\rm H}^2 - 2F_{\rm NH}F_{\rm H}\cos\psi, \qquad (2.4.4.27)$$

where $\psi = \alpha_{\rm NH} - \alpha_{\rm H}$. Using arguments similar to those used in deriving (2.4.3.5), we obtain

$$\sin\psi = [F_{\rm NH}^2(+) - F_{\rm NH}^2(-)]/4F_{\rm NH}F_{\rm H}''. \qquad (2.4.4.28)$$

If $F_{\rm NH}$ is considered to be equal to $[F_{\rm NH}(+) + F_{\rm NH}(-)]/2$, we obtain from (2.4.4.28)

$$F_{\rm NH}(+) - F_{\rm NH}(-) = 2F_{\rm H}''\sin\psi. \qquad (2.4.4.29)$$

We obtain what may be called $\psi_{\rm iso}$ if the magnitude of ψ is determined from (2.4.4.26) and the quadrant from (2.4.4.28). Similarly, we obtain $\psi_{\rm ano}$ if the magnitude of ψ is determined from (2.4.4.28) and the quadrant from (2.4.4.26). Ideally, $\psi_{\rm iso}$ and $\psi_{\rm ano}$ should have the same value and the difference between them is a measure of the errors in the data. $F_{\rm N}$ obtained from (2.4.4.27) using $\psi_{\rm ano}$ may be considered as its calculated value ($F_{\rm Ncal}$). Then, assuming all errors to lie in $F_{\rm N}$, we may write

$$E^2 = \sum_n (F_{\rm N} - F_{\rm Ncal})^2/n. \qquad (2.4.4.30)$$

Similarly, the calculated anomalous difference ($\Delta H_{\rm cal}$) may be evaluated from (2.4.4.29) using $\psi_{\rm iso}$. Then

$$E'^2 = \sum_n [|F_{\rm NH}(+) - F_{\rm NH}(-)| - \Delta H_{\rm cal}]^2/n. \qquad (2.4.4.31)$$

If all errors are assumed to reside in $F_{\rm H}$, E can be evaluated in yet another way using the expression

$$E^2 = \sum_n (F_{\rm HLE} - F_{\rm H})^2/n. \qquad (2.4.4.32)$$

2.4.4.7. Suggested modifications to Blow and Crick formulation and the inclusion of phase information from other sources

Modifications to the Blow and Crick procedure of phase evaluation have been suggested by several workers, although none represent a fundamental departure from the essential features of their formulation. In one of the modifications (Cullis *et al.*, 1961*a*; Ashida, 1976), all E_i's are assumed to be the same, but the lack-of-closure error $\xi_{{\rm H}i}$ for the ith derivative is measured as the distance from the mean of all intersections between phase circles to the point of intersection of the phase circle of that derivative with the phase circle of the native protein. Alternatively, individual values of E_i are retained, but the lack of closure is measured from the weighted mean of all intersections (Ashida, 1976). This is obviously designed to undo the effects of the unduly high weight given to $F_{\rm N}$ in the Blow and Crick formulation. In another modification (Raiz & Andreeva, 1970; Einstein, 1977), suggested for the same purpose, the $F_{\rm N}$ and $F_{{\rm NH}i}$ circles are treated as circular bands, the width of each band being related to the error in the appropriate structure factor. A comprehensive set of modifications suggested by Green (1979) treats different types of errors separately. In particular, errors arising from imperfect isomorphism are treated in a comprehensive manner.

Although the isomorphous replacement method still remains the method of choice for the *ab initio* determination of protein structures, additional items of phase information from other sources are increasingly being used to replace, supplement, or extend the information obtained through the application of the isomorphous replacement. Methods have been developed for the routine refinement of protein structures (Watenpaugh *et al.*, 1973; Huber *et al.*, 1974; Sussman *et al.*, 1977; Jack & Levitt, 1978; Isaacs & Agarwal, 1978; Hendrickson & Konnert, 1980) and they provide a rich source of phase information. However, the nature of the problem and the inherent limitations of the Fourier technique are such that the possibility of refinement yielding misleading results exists (Vijayan, 1980*a*,*b*). It is therefore sometimes desirable to combine the phases obtained during refinement with the original isomorphous replacement phases. The other sources of phase information include molecular replacement (see Chapter 2.3), direct methods (Hendrickson & Karle, 1973; Sayre, 1974; de Rango *et al.*, 1975; see also Chapter 2.2) and different types of electron-density modifications (Hoppe & Gassmann, 1968; Collins, 1975; Schevitz *et al.*, 1981; Bhat & Blow, 1982; Agard & Stroud, 1982; Cannillo *et al.*, 1983; Raghavan & Tulinsky, 1979; Wang, 1985).

The problem of combining isomorphous replacement phases with those obtained by other methods was first addressed by Rossmann & Blow (1961). The problem was subsequently examined by Hendrickson & Lattman (1970) and their method, which involves a modification of the Blow and Crick formulation, is perhaps the most widely used for combining phase information from different sources.

The Blow and Crick procedure is based on an assumed Gaussian 'lumped' error in $F_{{\rm NH}i}$ which leads to a lack of closure, $\xi_{{\rm H}i}(\alpha)$, in $F_{{\rm NH}i}$ defined by (2.4.4.20). Hendrickson and Lattman make an equally legitimate assumption that the lumped error, again assumed to be Gaussian, is associated with $F_{{\rm NH}i}^2$. Then, as in (2.4.4.20), we have

$$\xi_{{\rm H}i}''(\alpha) = F_{{\rm NH}i}^2 - D_{{\rm H}i}^2(\alpha), \qquad (2.4.4.33)$$

where $\xi_{{\rm H}i}''(\alpha)$ is the lack of closure associated with $F_{{\rm NH}i}^2$ for an assumed protein phase angle α. Then the probability for α being the correct phase angle can be expressed as

$$P_i(\alpha) = N_i \exp[-\xi_{{\rm H}i}''^2(\alpha)/2E_i''^2], \qquad (2.4.4.34)$$

where E_i'' is the r.m.s. error in $F_{{\rm NH}i}^2$, which can be evaluated using methods similar to those employed for evaluating E_i. Hendrickson and Lattman have shown that the exponent in the probability expression (2.4.4.34) can be readily expressed as a linear combination of five terms in the following manner.

$$-\xi_{{\rm H}i}''^2(\alpha)/2E_i''^2 = K_i + A_i\cos\alpha + B_i\sin\alpha + C_i\cos 2\alpha + D_i\sin 2\alpha, \qquad (2.4.4.35)$$

where K_i, A_i, B_i, C_i and D_i are constants dependent on $F_{\rm N}$, $F_{{\rm H}i}$, $F_{{\rm NH}i}$ and E_i''. Thus, five constants are enough to store the complete probability distribution of any reflection. Expressions for the five constants have been derived for phase information from anomalous scattering, tangent formula, partial structure and molecular replacement. The combination of the phase information from all sources can then be achieved by simply taking the total value of each constant. Thus, the total probability of the protein phase angle being α is given by

$$P(\alpha) = \prod P_s(\alpha) = N \exp\Bigg(\sum_s K_s + \sum_s A_s \cos\alpha + \sum_s B_s \sin\alpha + \sum_s C_s \cos 2\alpha + \sum_s D_s \sin 2\alpha \Bigg), \tag{2.4.4.36}$$

where K_s, A_s *etc.* are the constants appropriate for the sth source and N is the normalization constant.

2.4.4.8. Fourier representation of anomalous scatterers

It is often useful to have a Fourier representation of only the anomalous scatterers in a protein. The imaginary component of the electron-density distribution obviously provides such a representation. When the structure is known and $F_N(+)$ and $F_N(-)$ have been experimentally determined, Chacko & Srinivasan (1970) have shown that this representation is obtained in a Fourier synthesis with $i[\mathbf{F}_N(+) + \mathbf{F}_N^*(-)]/2$ as coefficients, where $\mathbf{F}_N^*(-)$, whose magnitude is $F_N(-)$, is the complex conjugate of $\mathbf{F}_N(+)$. They also indicated a method for calculating the phase angles of $\mathbf{F}_N(+)$ and $\mathbf{F}_N^*(-)$. It has been shown (Hendrickson & Sheriff, 1987) that the Bijvoet-difference Fourier synthesis proposed earlier by Kraut (1968) is an approximation of the true imaginary component of the electron density. The imaginary synthesis can be useful in identifying minor anomalous-scattering centres when the major centres are known and also in providing an independent check on the locations of anomalous scatterers and in distinguishing between anomalous scatterers with nearly equal atomic numbers (Sheriff & Hendrickson, 1987; Kitagawa *et al.*, 1987).

2.4.5. Anomalous scattering of neutrons and synchrotron radiation. The multiwavelength method

The multiwavelength anomalous-scattering method (Ramaseshan, 1982) relies on the variation of dispersion-correction terms as a function of the wavelength used. The success of the method therefore depends upon the size of the correction terms and the availability of incident beams of comparable intensities at different appropriate wavelengths. Thus, although this method was used as early as 1957 (Ramaseshan *et al.*, 1957) as an aid to structure solution employing characteristic X-rays, it is, as outlined below, ideally suited in structural work employing neutrons and synchrotron radiation. In principle, γ-radiation can also be used for phase determination (Raghavan, 1961; Moon, 1961) as the anomalous-scattering effects in γ-ray scattering could be very large; the wavelength is also easily tunable. However, the intensity obtainable for γ-rays is several orders lower than that obtainable from X-ray and neutron sources, and hence γ-ray anomalous scattering is of hardly any practical value in structural analysis.

2.4.5.1. Neutron anomalous scattering

Apart from the limitations introduced by experimental factors, such as the need for large crystals and the comparatively low flux of neutron beams, there are two fundamental reasons why neutrons are less suitable than X-rays for the *ab initio* determination of crystal structures. First, the neutron scattering lengths of different nuclei have comparable magnitudes whereas the atomic form factors for X-rays vary by two orders of magnitude. Therefore, Patterson techniques and the related heavy-atom method are much less suitable for use with neutron diffraction data than with X-ray data. Secondly, neutron scattering lengths could be positive or negative and hence, in general, the positivity criterion (Karle & Hauptman, 1950) or the squarability criterion (Sayre, 1952) does not hold good for nuclear density. Therefore, the direct methods based on these criteria are not strictly applicable to structure analysis using neutron data, although it has been demonstrated that these methods could be successfully used in favourable situations in neutron crystallography (Sikka, 1969). The anomalous-scattering method is, however, in principle more powerful in the neutron case than in the X-ray case for *ab initio* structure determination.

Thermal neutrons are scattered anomalously at appropriate wavelengths by several nuclei. In a manner analogous to (2.4.3.1), the neutron scattering length of these nuclei can be written as

$$b_0 + b' + ib'' = b + ib''. \tag{2.4.5.1}$$

The correction terms b' and b'' are strongly wavelength-dependent. In favourable cases, b'/b_0 and b''/b_0 can be of the order of 10 whereas they are small fractions in X-ray anomalous scattering. In view of this pronounced anomalous effect in neutron scattering, Ramaseshan (1966) suggested that it could be used for structure solution. Subsequently, Singh & Ramaseshan (1968) proposed a two-wavelength method for unique structure analysis using neutron diffraction. The first part of the method is the determination of the positions of the anomalous scatterers from the estimated values of F_Q. The method employed for estimating F_Q is analogous to that using (2.4.4.9) except that data collected at two appropriate wavelengths are used instead of those from two isomorphous crystals. The second stage of the two-wavelength method involves phase evaluation. Referring to Fig. 2.4.3.2 and in a manner analogous to (2.4.3.5), we have

$$\sin\psi_1 = \frac{F_{N1}^2(+) - F_{N1}^2(-)}{4F_{N1}F_{Q1}''}, \tag{2.4.5.2}$$

where $\psi = \alpha_N - \alpha_Q$ and subscript 1 refers to data collected at wavelength $\lambda 1$. Singh and Ramaseshan showed that $\cos\psi_1$ can also be determined when data are available at wavelength $\lambda 1$ and $\lambda 2$. We may define

$$F_m^2 = [F_N^2(+) + F_N^2(-)]/2 \tag{2.4.5.3}$$

and we have from (2.4.3.3), (2.4.3.4) and (2.4.5.3)

$$F_N = (F_m^2 - F_Q''^2)^{1/2}. \tag{2.4.5.4}$$

Then,

$$\cos\psi_1 = \frac{F_{m1}^2 - F_{m2}^2 - [(b_1^2 + b_1''^2) - (b_2^2 + b_2''^2)]x^2}{2(b_1 - b_2)F_{N1}x} + \frac{F_{Q1}}{F_{N1}}, \tag{2.4.5.5}$$

where x is the magnitude of the temperature-corrected geometrical part of $\mathbf{F}_Q$. ψ_1 and hence α_{N1} can be calculated using (2.4.5.2) and (2.4.5.5). α_{N2} can also be obtained in a similar manner.

During the decade that followed Ramaseshan's suggestion, neutron anomalous scattering was used to solve half a dozen crystal structures, employing the multiple-wavelength methods as well as the methods developed for structure determination using X-ray anomalous scattering (Koetzle & Hamilton, 1975; Sikka & Rajagopal, 1975; Flook *et al.*, 1977). It has also been demonstrated that measurable Bijvoet differences could be obtained, in favourable situations, in neutron diffraction patterns from protein crystals (Schoenborn, 1975). However, despite the early promise held by neutron anomalous scattering, the method has

not been as successful as might have been hoped. In addition to the need for large crystals, the main problem with using this method appears to be the time and expense involved in data collection (Koetzle & Hamilton, 1975).

2.4.5.2. Anomalous scattering of synchrotron radiation

The most significant development in recent years in relation to anomalous scattering of X-rays has been the advent of synchrotron radiation (Helliwell, 1984). The advantage of using synchrotron radiation for making anomalous-scattering measurements essentially arises out of the tunability of the wavelength. Unlike the characteristic radiation from conventional X-ray sources, synchrotron radiation has a smooth spectrum and the wavelength to be used can be finely selected. Accurate measurements have shown that values in the neighbourhood of 30 electrons could be obtained in favourable cases for f' and f'' (Templeton, Templeton, Phillips & Hodgson, 1980; Templeton, Templeton & Phizackerley, 1980; Templeton *et al.*, 1982). Schemes for the optimization of the wavelengths to be used have also been suggested (Narayan & Ramaseshan, 1981). Interestingly, the anomalous differences obtainable using synchrotron radiation are comparable in magnitude to the isomorphous differences normally encountered in protein crystallography. Thus, the use of anomalous scattering at several wavelengths would obviously eliminate the need for employing many heavy-atom derivatives. The application of anomalous scattering of synchrotron radiation for macromolecular structure analysis began to yield encouraging results in the 1980s (Helliwell, 1985). Intensity measurements from macromolecular X-ray diffraction patterns using synchrotron radiation at first relied primarily upon oscillation photography (Arndt & Wonacott, 1977). This method is not particularly suitable for accurately evaluating anomalous differences. Much higher levels of accuracy began to be achieved with the use of position-sensitive detectors (Arndt, 1986). Anomalous scattering, in combination with such detectors, has developed into a major tool in macromolecular crystallography (see *IT* F, 2001).

One of us (MV) acknowledges the support of the Department of Science & Technology, India.

References

Adams, M. J. (1968). DPhil thesis, Oxford University, England.

Agard, D. A. & Stroud, R. M. (1982). *α-Bungarotoxin structure revealed by a rapid method for averaging electron density of non-crystallographically translationally related molecules. Acta Cryst.* A**38**, 186–194.

Arndt, U. W. (1986). *X-ray position-sensitive detectors. J. Appl. Cryst.* **19**, 145–163.

Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

Ashida, T. (1976). *Some remarks on the phase angle determination by the isomorphous replacement method*. In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 282–284. Copenhagen: Munksgaard.

Bhat, T. N. & Blow, D. M. (1982). *A density-modification method for improvement of poorly resolved protein electron-density maps. Acta Cryst.* A**38**, 21–29.

Bijvoet, J. M. (1949). *Phase determination in direct Fourier-synthesis of crystal structures. Proc. K. Ned. Akad. Wet. (B)*, **52**, 313–314.

Bijvoet, J. M. (1954). *Structure of optically active compounds in the solid state. Nature (London),* **173**, 888–891.

Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. (1951). *Determination of the absolute configuration of optically active compounds by means of X-rays. Nature (London)*, **168**, 271–272.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Blow, D. M. & Matthews, B. W. (1973). *Parameter refinement in the multiple isomorphous-replacement method. Acta Cryst.* A**29**, 56–62.

Blow, D. M. & Rossmann, M. G. (1961). *The single isomorphous replacement method. Acta Cryst.* **14**, 1195–1202.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. London: Academic Press.

Bokhoven, C., Schoone, J. C. & Bijvoet, J. M. (1951). *The Fourier synthesis of the crystal structure of strychnine sulphate pentahydrate. Acta Cryst.* **4**, 275–280.

Bradley, A. J. & Rodgers, J. W. (1934). *The crystal structure of the Heusler alloys. Proc. R. Soc. London Ser. A*, **144**, 340–359.

Cannillo, E., Oberti, R. & Ungaretti, L. (1983). *Phase extension and refinement by density modification in protein crystallography. Acta Cryst.* A**39**, 68–74.

Chacko, K. K. & Srinivasan, R. (1970). *On the Fourier refinement of anomalous dispersion corrections in X-ray diffraction data. Z. Kristallogr.* **131**, 88–94.

Collins, D. M. (1975). *Efficiency in Fourier phase refinement for protein crystal structures. Acta Cryst.* A**31**, 388–389.

Cork, J. M. (1927). *The crystal structure of some of the alums. Philos. Mag.* **4**, 688–698.

Coster, D., Knol, K. S. & Prins, J. A. (1930). *Unterschiede in der Intensität der Röntgenstrahlenreflexion an den beiden 111-Flachen der Zinkblende. Z. Phys.* **63**, 345–369.

Cromer, D. T. (1965). *Anomalous dispersion corrections computed from self-consistent field relativistic Dirac–Slater wave functions. Acta Cryst.* **18**, 17–23.

Cruickshank, D. W. J. & McDonald, W. S. (1967). *Parameter errors in polar space groups caused by neglect of anomalous scattering. Acta Cryst.* **23**, 9–11.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1961*a*). *The structure of haemoglobin. VIII. A three-dimensional Fourier synthesis at 5.5 Å resolution: determination of the phase angles. Proc. R. Soc. London Ser. A*, **265**, 15–38.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1961*b*). *The structure of haemoglobin. IX. A three-dimensional Fourier synthesis at 5.5 Å resolution: description of the structure. Proc. R. Soc. London Ser. A*, **265**, 161–187.

Dale, D., Hodgkin, D. C. & Venkatesan, K. (1963). *The determination of the crystal structure of factor V 1a*. In *Crystallography and Crystal Perfection*, edited by G. N. Ramachandran, pp. 237–242. New York, London: Academic Press.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. Acta Cryst.* **14**, 1188–1195.

Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. (1968). *A least-squares refinement method for isomorphous replacement. Acta Cryst.* B**24**, 997–1003.

Dodson, E., Evans, P. & French, S. (1975). *The use of anomalous scattering in refining heavy atom parameters in proteins.* In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 423–436. Copenhagen: Munksgaard.

Dodson, E. & Vijayan, M. (1971). *The determination and refinement of heavy-atom parameters in protein heavy-atom derivatives. Some model calculations using acentric reflexions. Acta Cryst.* B**27**, 2402–2411.

Einstein, J. E. (1977). *An improved method for combining isomorphous replacement and anomalous scattering diffraction data for macromolecular crystals. Acta Cryst.* A**33**, 75–85.

Flook, R. J., Freeman, H. C. & Scudder, M. L. (1977). *An X-ray and neutron diffraction study of aqua(L-glutamato) cadmium(II) hydrate. Acta Cryst.* B**33**, 801–809.

Gilli, G. & Cruickshank, D. W. J. (1973). *Effect of neglect of dispersion in centrosymmetric structures: results for OsO_4. Acta Cryst.* B**29**, 1983–1985.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. Proc. R. Soc. London Ser A*, **225**, 287–307.

Green, E. A. (1979). *A new statistical model for describing errors in isomorphous replacement data: the case of one derivative. Acta Cryst.* A**35**, 351–359.

Harker, D. (1956). *The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Cryst.* **9**, 1–9.

Helliwell, J. R. (1984). *Synchrotron X-radiation protein crystallography: instrumentation, methods and applications. Rep. Prog. Phys.* **47**, 1403–1497.

Helliwell, J. R. (1985). *Protein crystallography with synchrotron radiation. J. Mol. Struct.* **130**, 63–91.

Hendrickson, W. A. (1979). *Phase information from anomalous-scattering measurements. Acta Cryst.* A**35**, 245–247.

Hendrickson, W. A. & Karle, J. (1973). *Carp muscle calcium-binding protein. III. Phase refinement using the tangent formula. J. Biol. Chem.* **248**, 3327–3334.

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of stereochemical information into crystallographic refinement.* In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

Hendrickson, W. A. & Lattman, E. E. (1970). *Representation of phase probability distributions in simplified combinations of independent phase information. Acta Cryst.* B**26**, 136–143.

Hendrickson, W. A. & Sheriff, S. (1987). *General density function corresponding to X-ray diffraction with anomalous scattering included. Acta Cryst.* A**43**, 121–125.

Hendrickson, W. A. & Teeter, M. M. (1981). *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London)*, **290**, 107–113.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). *Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. J. Mol. Biol.* **89**, 73–101.

Ibers, J. A. & Hamilton, W. C. (1964). *Dispersion corrections and crystal structure refinements. Acta Cryst.* **17**, 781–782.

International Tables for Crystallography (2001). Vol. F, *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold. Dordrecht: Kluwer Academic Publishers.

International Tables for X-ray Crystallography (1974). Vol. IV, pp. 148–151. Birmingham: Kynoch Press.

Isaacs, N. W. & Agarwal, R. C. (1978). *Experience with fast Fourier least squares in the refinement of the crystal structure of rhombohedral 2-zinc insulin at 1.5 Å resolution. Acta Cryst.* A**34**, 782–791.

Jack, A. & Levitt, M. (1978). *Refinement of large structures by simultaneous minimization of energy and R factor. Acta Cryst.* A**34**, 931–935.

Karle, J. & Hauptman, H. (1950). *The phases and magnitudes of the structure factors. Acta Cryst.* **3**, 181–187.

Kartha, G. (1961). *Isomorphous replacement in non-centrosymmetric structures. Acta Cryst.* **14**, 680–686.

Kartha, G. (1965). *Combination of multiple isomorphous replacement and anomalous dispersion data for protein structure determination. III. Refinement of heavy atom positions by the least-squares method. Acta Cryst.* **19**, 883–885.

Kartha, G. (1975). *Application of anomalous scattering studies in protein structure analysis.* In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 363–392. Copenhagen: Munksgaard.

Kartha, G. (1976). *Protein phase evaluation: multiple isomorphous series and anomalous scattering methods.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 269–281. Copenhagen: Munksgaard.

Kartha, G. & Parthasarathy, R. (1965). *Combination of multiple isomorphous replacement and anomalous dispersion data for protein structure determination. I. Determination of heavy-atom positions in protein derivatives. Acta Cryst.* **18**, 745–749.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Phillips, D. C. & Shore, V. C. (1960). *Structure of myoglobin. A three-dimensional Fourier synthesis at 2 Å resolution. Nature (London)*, **185**, 422–427.

Kitagawa, Y., Tanaka, N., Hata, Y., Katsube, Y. & Satow, Y. (1987). *Distinction between Cu^{2+} and Zn^{2+} ions in a crystal of spinach superoxide dismutase by use of anomalous dispersion and tuneable synchrotron radiation. Acta Cryst.* B**43**, 272–275.

Koetzle, T. F. & Hamilton, W. C. (1975). *Neutron diffraction study of $NaSmEDTA.8H_2O$: an evaluation of methods of phase determination based on three-wavelength anomalous dispersion data.* In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 489–502. Copenhagen: Munksgaard.

Kraut, J. (1968). *Bijvoet-difference Fourier function. J. Mol. Biol.* **35**, 511–512.

Kraut, J., Sieker, L. C., High, D. F. & Freer, S. T. (1962). *Chymotrypsin: a three-dimensional Fourier synthesis at 5 Å resolution. Proc. Natl Acad. Sci. USA*, **48**, 1417–1424.

McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: John Wiley.

Mark, H. & Szillard, L. (1925). *Ein einfacher versuch zur auffinclung eines selectiven effecktes bei der zerstrenung von Röntgenstrahlen. Z. Phys.* **33**, 688–691.

Matthews, B. W. (1966). *The determination of the position of anomalously scattering heavy atom groups in protein crystals. Acta Cryst.* **20**, 230–239.

Moon, P. B. (1961). *Developments in gamma-ray optics. Nature (London)*, **185**, 427–429.

Muirhead, H., Cox, J. M., Mazzarella, L. & Perutz, M. F. (1967). *Structure and function of haemoglobin. III. A three-dimensional Fourier synthesis of human deoxyhaemoglobin at 5.5 Å resolution. J. Mol. Biol.* **28**, 156–177.

Narayan, R. & Ramaseshan, S. (1981). *Optimum choice of wavelengths in the anomalous scattering technique with synchrotron radiation. Acta Cryst.* A**37**, 636–641.

Navia, M. A. & Sigler, P. B. (1974). *The application of direct methods to the analysis of heavy-atom derivatives. Acta Cryst.* A**30**, 706–712.

North, A. C. T. (1965). *The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. Acta Cryst.* **18**, 212–216.

Okaya, Y. & Pepinsky, R. (1956). *New formulations and solution of the phase problem in X-ray analysis for noncentric crystals containing anomalous scatterers. Phys. Rev.* **103**, 1645–1657.

Parthasarathy, S. (1967). *Expectation value of the Bijvoet ratio. Acta Cryst.* **22**, 98–103.

Parthasarathy, S. & Srinivasan, R. (1964). *The probability distribution of Bijvoet differences. Acta Cryst.* **17**, 1400–1407.

Patterson, A. L. (1963). *Treatment of anomalous dispersion in X-ray diffraction data. Acta Cryst.* **16**, 1255–1256.

Peerdeman, A. F. & Bijvoet, J. M. (1956). *The indexing of reflexions in investigations involving the use of the anomalous scattering effect. Acta Cryst.* **9**, 1012–1015.

Peerdeman, A. F., van Bommel, A. J. & Bijvoet, J. M. (1951). *Determination of the absolute configuration of optically active compounds by means of X-rays. Proc. K. Ned. Akad. Wet. (B)*, **54**, 16–19.

Pepinsky, R., Okaya, Y. & Takeuchi, Y. (1957). *Theory and application of the $P_s(\mathbf{u})$ function and anomalous dispersion in direct determination of structures and absolute configuration in non-centric crystals. Acta Cryst.* **10**, 756.

Phillips, D. C. (1966). *Advances in protein crystallography.* In *Advances in Structure Research by Diffraction Methods*, Vol. 2, edited by R. Brill & R. Mason, pp. 75–140. New York, London: Interscience.

Raghavan, N. V. & Tulinsky, A. (1979). *The structure of α-chymotrypsin. II. Fourier phase refinement and extension of the dimeric structure at 1.8 Å resolution by density modification. Acta Cryst.* B**35**, 1776–1785.

Raghavan, R. S. (1961). *On the possibility of a new phase-determining method applying the Mössbauer effect. Proc. Indian Acad. Sci. Sect. A*, **53**, 265–271.

Raiz, V. Sh. & Andreeva, N. S. (1970). *Determining the coefficients of the Fourier series of the electron density function of protein crystals. Sov. Phys. Crystallogr.* **15**, 206–210.

Ramachandran, G. N. & Raman, S. (1956). *A new method for the structure analysis of non-centrosymmetric crystals. Curr. Sci.* **25**, 348–351.

Ramachandran, G. N. & Raman, S. (1959). *Syntheses for the deconvolution of the Patterson function. Part I. General principles. Acta Cryst.* **12**, 957–964.

Ramachandran, G. N. & Srinivasan, R. (1970). *Fourier Methods in Crystallography*. New York: Wiley–Interscience.

Ramaseshan, S. (1963). *The use of anomalous scattering in crystal structure analysis.* In *Advanced Methods of Crystallography*, edited by G. N. Ramachandran, pp. 67–95. London, New York: Academic Press.

Ramaseshan, S. (1966). *The use of anomalous scattering of neutrons in the solution of crystal structures containing large molecules. Curr. Sci.* **35**, 87–91.

Ramaseshan, S. (1982). *Use of the multi-wavelength anomalous scattering method in crystallography.* In *Computational Crystallography*, edited by D. Sayre, pp. 201–208. Oxford: Clarendon Press.

Ramaseshan, S. & Abrahams, S. C. (1975). *Anomalous Scattering*. Copenhagen: Munksgaard.

Ramaseshan, S. & Venkatesan, K. (1957). *The use of anomalous scattering without phase change in crystal structure analysis. Curr. Sci.* **26**, 352–353.

Ramaseshan, S., Venkatesan, K. & Mani, N. V. (1957). *The use of anomalous scattering for the determination of crystal structures – $KMnO_4$. Proc. Indian Acad. Sci.* **46**, 95–111.

Rango, C. de, Mauguen, Y. & Tsoucaris, G. (1975). *Use of high-order probability laws in phase refinement and extension of protein structures. Acta Cryst.* A**31**, 227–233.

Robertson, J. M. (1936). *An X-ray study of the phthalocyanines. Part II. Quantitative structure determination of the metal-free compound. J. Chem. Soc.* pp. 1195–1209.

Robertson, J. M. & Woodward, I. (1937). *An X-ray study of the phthalocyanines. Part III. Quantitative structure determination of nickel phthalocyanine. J. Chem. Soc.* pp. 219–230.

Rossmann, M. G. (1960). *The accurate determination of the position and shape of heavy-atom replacement groups in proteins. Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961). *The position of anomalous scatterers in protein crystals. Acta Cryst.* **14**, 383–388.

Rossmann, M. G. & Blow, D. M. (1961). *The refinement of structures partially determined by the isomorphous replacement method. Acta Cryst.* **14**, 641–647.

Sayre, D. (1952). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Sayre, D. (1974). *Least-squares phase refinement. II. High-resolution phasing of a small protein. Acta Cryst.* A**30**, 180–184.

Schevitz, R. W., Podjarny, A. D., Zwick, M., Hughes, J. J. & Sigler, P. B. (1981). *Improving and extending the phases of medium- and low-resolution macromolecular structure factors by density modification. Acta Cryst.* A**37**, 669–677.

Schoenborn, B. P. (1975). *Phasing of neutron protein data by anomalous dispersion*. In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 407–416, Copenhagen: Munksgaard.

Sheriff, S. & Hendrickson, W. A. (1987). *Location of iron and sulfur atoms in myohemerythrin from anomalous-scattering measurements. Acta Cryst.* B**43**, 209–212.

Sikka, S. K. (1969). *On the application of the symbolic addition procedure in neutron diffraction structure determination. Acta Cryst.* A**25**, 539–543.

Sikka, S. K. & Rajagopal, H. (1975). *Application of neutron anomalous dispersion in the structure determination of cadmium tartrate pentahydrate*. In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 503–514. Copenhagen: Munksgaard.

Singh, A. K. & Ramaseshan, S. (1966). *The determination of heavy atom positions in protein derivatives. Acta Cryst.* **21**, 279–280.

Singh, A. K. & Ramaseshan, S. (1968). *The use of neutron anomalous scattering in crystal structure analysis. I. Non-centrosymmetric structures. Acta Cryst.* B**24**, 35–39.

Srinivasan, R. (1972). *Applications of X-ray anomalous scattering in structural studies*. In *Advances in Structure Research by Diffraction Methods*, Vol. 4, edited by W. Hoppe & R. Mason, pp. 105–197. Braunschweig: Freidr. Vieweg & Sohn; and Oxford: Pergamon Press.

Srinivasan, R. & Chacko, K. K. (1970). *On the determination of phases of a noncentrosymmetric crystal by the anomalous dispersion method. Z. Kristallogr.* **131**, 29–39.

Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. (1977). *A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. Acta Cryst.* A**33**, 800–804.

Templeton, D. H., Templeton, L. K., Phillips, J. C. & Hodgson, K. O. (1980). *Anomalous scattering of X-rays by cesium and cobalt measured with synchrotron radiation. Acta Cryst.* A**36**, 436–442.

Templeton, L. K., Templeton, D. H. & Phizackerley, R. P. (1980). *L_3-edge anomalous scattering of X-rays by praseodymium and samarium. J. Am. Chem. Soc.* **102**, 1185–1186.

Templeton, L. K., Templeton, D. H., Phizackerley, R. P. & Hodgson, K. O. (1982). *L_3-edge anomalous scattering by gadolinium and samarium measured at high resolution with synchrotron radiation. Acta Cryst.* A**38**, 74–78.

Terwilliger, T. C. & Eisenberg, D. (1983). *Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions. Acta Cryst.* A**39**, 813–817.

Ueki, T., Zalkin, A. & Templeton, D. H. (1966). *Crystal structure of thorium nitrate pentahydrate by X-ray diffraction. Acta Cryst.* **20**, 836–841.

Vijayan, M. (1980*a*). *On the Fourier refinement of protein structures. Acta Cryst.* A**36**, 295–298.

Vijayan, M. (1980*b*). *Phase evaluation and some aspects of the Fourier refinement of macromolecules*. In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 19.01–19.25. Bangalore: Indian Academy of Sciences.

Vijayan, M. (1981). *X-ray analysis of 2Zn insulin: some crystallographic problems*. In *Structural Studies on Molecules of Biological Interest*, edited by G. Dodson, J. P. Glusker & D. Sayre, pp. 260–273. Oxford: Clarendon Press.

Vijayan, M. (1987). *Anomalous scattering methods.* In *Direct Methods, Macromolecular Crystallography and Crystallographic Statistics*, edited by H. Schenk, A. J. C. Wilson & S. Parthasarathy, pp. 121–139. Singapore: World Scientific Publishing Co. Pte. Ltd.

Vos, A. (1975). *Anomalous scattering and chemical crystallography*. In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 307–317. Copenhagen: Munksgaard.

Wang, B. C. (1985). *Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol.* **115**, 90–112.

Watenpaugh, K. D., Sieker, L. C., Herriot, J. R. & Jensen, L. H. (1973). *Refinement of the model of a protein: rubredoxin at 1.5 Å resolution. Acta Cryst.* B**29**, 943–956.

Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1975). *Anomalous scattering in protein structure analysis*. In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 393–405. Copenhagen: Munksgaard.

Wilson, A. J. C. (1942). *Determination of absolute from relative X-ray intensity data. Nature (London)*, **150**, 151–152.

Wilson, A. J. C. (1975). *Effect of neglect of dispersion on apparent scale and temperature parameters*. In *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 325–332. Copenhagen: Munksgaard.

Zachariasen, W. H. (1965). *Dispersion in quartz. Acta Cryst.* **18**, 714–716.

International Tables for Crystallography (2010). Vol. B, Chapter 2.5, pp. 297–402.

2.5. Electron diffraction and electron microscopy in structure determination

By J. M. Cowley,† J. C. H. Spence, M. Tanaka, B. K. Vainshtein,† B. B. Zvyagin,† P. A. Penczek and D. L. Dorset

2.5.1. Foreword

By J. M. Cowley and J. C. H. Spence

Given that electrons have wave properties and the wavelengths lie in a suitable range, the diffraction of electrons by matter is completely analogous to the diffraction of X-rays. While for X-rays the scattering function is the electron-density distribution, for electrons it is the potential distribution which is similarly peaked at the atomic sites. Hence, in principle, electron diffraction may be used as the basis for crystal structure determination. In practice it is used much less widely than X-ray diffraction for the determination of crystal structures but is receiving increasing attention as a means for obtaining structural information not readily accessible with X-ray- or neutron-diffraction techniques.

Electrons having wavelengths comparable with those of the X-rays commonly used in diffraction experiments have energies of the order of 100 eV. For such electrons, the interactions with matter are so strong that they can penetrate only a few layers of atoms on the surfaces of solids. They are used extensively for the study of surface structures by low-energy electron diffraction (LEED) and associated techniques. These techniques are not covered in this series of volumes, which include the principles and practice of only those diffraction and imaging techniques making use of high-energy electrons, having energies in the range of 20 keV to 1 MeV or more, in transmission through thin specimens.

For the most commonly used energy ranges of high-energy electrons, 100 to 400 keV, the wavelengths are about 50 times smaller than for X-rays. Hence the scattering angles are much smaller, of the order of 10^{-2} rad, the recording geometry is relatively simple and the diffraction pattern represents, to a useful first approximation, a planar section of reciprocal space. Extinction distances are hundreds of ångstroms, which, when combined with typical lattice spacings, produces rocking-curve widths which are, unlike the X-ray case, a significant fraction of the Bragg angle.

The elastic scattering of electrons by atoms is several orders of magnitude greater than for X-rays. This fact has profound consequences, which in some cases are highly favourable and in other cases are serious hindrances to structure analysis work. On the one hand it implies that electron-diffraction patterns can be obtained from very small single-crystal regions having thicknesses equal to only a few layers of atoms and, with recently developed techniques, having diameters equivalent to only a few interatomic distances. Hence single-crystal patterns can be obtained from microcrystalline phases.

However, the strong scattering of electrons implies that the simple kinematical single-scattering approximation, on which most X-ray diffraction structure analysis is based, fails for electrons except for very thin crystals composed of light-atom materials. Strong dynamical diffraction effects occur for crystals which may be 100 Å thick, or less for heavy-atom materials. As a consequence, the theory of dynamical diffraction for electrons has been well developed, particularly for the particular special diffracting conditions relevant to the transmission of fast electrons (see Chapter 5.2), and observations of dynamical diffraction effects are commonly made and quantitatively interpreted. The possibility has thus arisen of using the observation of dynamical diffraction effects as the basis for obtaining crystal structure information. The fact that dynamical diffraction is dependent on the relative phases of the diffracted waves then implies that relative phase information can be deduced from the diffraction intensities and the limitations of kinematical diffraction, such as Friedel's law, do not apply. The most immediately practicable method for making use of this possibility is convergent-beam electron diffraction (CBED) as described in Section 2.5.3.

A further important factor, determining the methods for observing electron diffraction, is that, being charged particles, electrons can be focused by electromagnetic lenses. Many of the resolution-limiting aberrations of cylindrical magnetic lenses have now been eliminated through the use of aberration-correction devices, so that for weakly scattering samples the resolution is limited to about 1 Å by electronic and mechanical instabilities. This is more than sufficient to distinguish the individual rows of atoms, parallel to the incident beam, in the principal orientations of most crystalline phases. Thus 'structure images' can be obtained, sometimes showing direct representation of projections of crystal structures [see *IT* C (2004), Section 4.3.8]. However, the complications of dynamical scattering and of the coherent imaging processes are such that the image intensities vary strongly with crystal thickness and tilt, and with the defocus or other parameters of the imaging system, making the interpretation of images difficult except in special circumstances. Fortunately, computer programs are readily available whereby image intensities can be calculated for model structures [see *IT* C (2004), Section 4.3.6]. Hence the means exist for deriving the projection of the structure if only by a process of trial and error and not, as would be desirable, from a direct interpretation of the observations.

The accuracy with which the projection of a structure can be deduced from an image, or series of images, improves as the resolution of the microscope improves but is not at all comparable with the accuracy attainable with X-ray diffraction methods. A particular virtue of high-resolution electron microscopy as a structural tool is that it may give information on individual small regions of the sample. Structures can be determined of 'phases' existing over distances of only a few unit cells and the defects and local disorders can be examined, one by one.

The observation of electron-diffraction patterns forms an essential part of the technique of structure imaging in high-resolution electron microscopy, because the diffraction patterns are used to align the crystals to appropriate axial orientations. More generally, for all electron microscopy of crystalline materials the image interpretation depends on knowledge of the diffraction conditions. Fortunately, the diffraction pattern and image of any specimen region can be obtained in rapid succession by a simple switching of lens currents. The ready comparison of the image and diffraction data has become an essential component of the electron microscopy of crystalline materials but has also been of fundamental importance for the development of electron-diffraction theory and techniques. The development of the nanodiffraction method in the field-emission scanning transmission electron microscope (STEM) has allowed microdiffraction patterns to be obtained from subnanometre-sized regions, and so has become the ideal tool for the structural analysis of the new microcrystalline phases important to nanoscience. The direct phasing of these coherent nanodiffraction patterns is an active field of research.

† Deceased.

Copyright © 2010 International Union of Crystallography

The individual specimen regions giving single-crystal electron-diffraction patterns are, with few exceptions, so small that they can be seen only by use of an electron microscope. Hence, historically, it was only after electron microscopes were commonly available that the direct correlations of diffraction intensities with crystal size and shape could be made, and a proper basis was available for the development of the adequate dynamical diffraction theory.

For the complete description of a diffraction pattern or image intensities obtained with electrons, it is necessary to include the effects of inelastic scattering as well as elastic scattering. In contrast to the X-ray diffraction case, the inelastic scattering does not produce just a broad and generally negligible background. The average energy loss for an inelastic scattering event is about 20 eV, which is small compared with the energy of about 100 keV for the incident electrons. The inelastically scattered electrons have a narrow angular distribution and are diffracted in much the same way as the incident or elastically scattered electrons in a crystal. They therefore produce a highly modulated contribution to the diffraction pattern, strongly peaked about the Bragg spot positions (see Chapter 4.3). Also, as a result of the inelastic scattering processes, including thermal diffuse scattering, an effective absorption function must be added in the calculation of intensities for elastically scattered electrons.

The inelastic scattering processes in themselves give information about the specimen in that they provide a measure of the excitations of both the valence-shell and the inner-shell electrons of the solid. The inner-shell electron excitations are characteristic of the type of atom, so that microanalysis of small volumes of specimen material (a few hundreds or thousands of atoms) may be achieved by detecting either the energy losses of the transmitted electrons or the emission of the characteristic X-ray [see *IT* C (2004), Section 4.3.4]. The development of the annular dark field (ADF) mode in STEM provides a favourable detector geometry for microanalysis, in which the forward scattered beam may be passed to an electron energy-loss spectrometer (EELS) for spectral analysis, while scattering at larger angles is collected to form a simultaneous scanning image. The arrangement is particularly efficient because, using a magnetic sector dispersive spectrometer, electrons of all energy losses may be detected simultaneously (parallel detection). Fine structure on the EELS absorption edges is analysed in a manner analogous to soft X-ray absorption spectroscopy, but with a spatial resolution of a few nanometres. The spectra are obtained from points in the corresponding ADF image which can be identified with subnanometre accuracy.

An adverse effect of the inelastic scattering processes, however, is that the transfer of energy to the specimen material results in radiation damage; this is a serious limitation of the application of electron-scattering methods to radiation-sensitive materials such as organic, biological and many inorganic compounds. The amount of radiation damage increases rapidly as the amount of information per unit volume, derived from the elastic scattering, is increased, *i.e.* as the microscope resolution is improved or as the specimen volume irradiated during a diffraction experiment is decreased. At the current limits of microscopic resolution, radiation damage is a significant factor even for the radiation-resistant materials such as semiconductors and alloys.

In the historical development of electron-diffraction techniques the progress has depended to an important extent on the level of understanding of the dynamical diffraction processes and this understanding has followed, to a considerable degree, from the availability of electron microscopes. For the first 20 years of the development, with few exceptions, the lack of a precise knowledge of the specimen morphology meant that diffraction intensities were influenced to an unpredictable degree by dynamical scattering and the impression grew that electron-diffraction intensities could not meaningfully be interpreted.

It was the group in the Soviet Union, led initially by Dr Z. G. Pinsker and later by Dr B. K. Vainshtein and others, which showed that patterns from thin layers of a powder of microcrystals could be interpreted reliably by use of the kinematical approximation. The averaging over crystal orientation reduced the dynamical diffraction effects to the extent that practical structure analysis was feasible. The development of the techniques of using films of crystallites having strongly preferred orientations, to give patterns somewhat analogous to the X-ray rotation patterns, provided the basis for the collection of three-dimensional diffraction data on which many structure analyses have been based [see Section 2.5.4 and *IT* C (2004), Section 4.3.5].

In recent years improvements in the techniques of specimen preparation and in the knowledge of the conditions under which dynamical diffraction effects become significant have allowed progress to be made with the use of high-energy electron-diffraction patterns from thin single crystals for crystal structure analysis. Particularly for crystals of light-atom materials, including biological and organic compounds, the methods of structure analysis developed for X-ray diffraction, including the direct methods (see Section 2.5.8), have been successfully applied in an increasing number of cases. Often it is possible to deduce some structural information from high-resolution electron-microscope images and this information may be combined with that from the diffraction intensities to assist the structure analysis process [see *IT* C (2004), Section 4.3.8.8].

The determination of crystal symmetry by use of CBED (Section 2.5.3) and the accurate determination of structure amplitudes by use of methods depending on the observation of dynamical diffraction effects [*IT* C (2004), Section 4.3.7] came later, after the information on morphologies of crystals, and the precision electron optics associated with electron microscopes, became available. This powerful convergent-beam microdiffraction method has now been widely adopted as the preferred method for space-group determination of microphases, quasicrystals, incommensurate, twinned and other imperfectly crystalline structures. Advantage is taken of the fact that multiple scattering preserves information on the absence of inversion symmetry, while the use of an electron probe which is smaller than a mosaic block allows extinction-free structure-factor measurements to be made. Finally, an enhanced sensitivity to ionicity is obtained from electron-diffraction measurements of structure factors by the very large difference between electron scattering factors for atoms and those for ions at small angles. This section by M. Tanaka replaces the corresponding section by the late P. Goodman in previous editions, which researchers may also find useful.

In spite of the problem of radiation damage, a great deal of progress has been made in the study of organic and biological materials by electron-scattering methods. In some respects these materials are very favourable because, with only light atoms present, the scattering from thin films can be treated using the kinematical approximation without serious error. Because of the problem of radiation damage, however, special techniques have been evolved to maximize the information on the required structural aspects with minimum irradiation of the specimen. Image-processing techniques have been evolved to take advantage of the redundancy of information from a periodic structure and the means have been devised for combining information from multiple images and diffraction data to reconstruct specimen structure in three dimensions. These techniques are outlined in Sections 2.5.5, 2.5.6 and 2.5.7. Section 2.5.6, written for the first and second editions by Boris Vainshtein, has been revised and extended for this third edition by Pawel Penczek. It deals with the general theory of three-dimensional reconstruction from projections and compares several popular methods. Section 2.5.7 describes the application of electron-microscope imaging to the structure analysis of proteins which cannot be crystallized,

and so addresses a crucial problem in structural biology. This is done by the remarkably successful method of single-particle image reconstruction, in which images of the same protein, lying in random orientations within a thin film of vitreous ice, are combined in the correct orientation to form a three-dimensional reconstructed charge-density map at nanometre or better resolution. The summation over many particles achieves the same radiation-damage-reduction effect as does crystallographic redundancy in protein crystallography. Finally, Section 2.5.8 describes experience with the application of numerical direct methods to the phase problem in electron diffraction. Although direct imaging 'solves' the phase problem, there are many practical problems in combining electron-microdiffraction intensities with corresponding high-resolution images of a structure over a large tilt range. In cases where multiple scattering can be minimized, some success has therefore been obtained using direct phasing methods, as reviewed in this section.

For most inorganic materials the complications of many-beam dynamical diffraction processes prevent the direct application of these techniques of image analysis, which depend on having a linear relationship between the image intensity and the value of the projected potential distribution of the sample. The smaller sensitivities to radiation damage can, to some extent, remove the need for the application of such methods by allowing direct visualization of structure with ultra-high-resolution images and the use of microdiffraction techniques.

2.5.2. Electron diffraction and electron microscopy[1]

By J. M. Cowley

2.5.2.1. Introduction

The contributions of electron scattering to the study of the structures of crystalline solids are many and diverse. This section will deal only with the scattering of high-energy electrons (in the energy range of 10^4 to 10^6 eV) in transmission through thin samples of crystalline solids and the derivation of information on crystal structures from diffraction patterns and high-resolution images. The range of wavelengths considered is from about 0.122 Å (12.2 pm) for 10 kV electrons to 0.0087 Å (0.87 pm) for 1 MeV electrons. Given that the scattering amplitudes of atoms for electrons have much the same form and variation with $(\sin\theta)/\lambda$ as for X-rays, it is apparent that the angular range for strong scattering of electrons will be of the order of 10^{-2} rad. Only under special circumstances, usually involving multiple elastic and inelastic scattering from very thick specimens, are scattering angles of more than 10^{-1} rad of importance.

The strength of the interaction of electrons with matter is greater than that of X-rays by two or three orders of magnitude. The single-scattering, first Born approximation fails significantly for scattering from single heavy atoms. Diffracted beams from single crystals may attain intensities comparable with that of the incident beam for crystal thicknesses of 10^2 Å, rather than 10^4 Å or more. It follows that electrons may be used for the study of very thin samples, and that dynamical scattering effects, or the coherent interaction of multiply scattered electron waves, will modify the diffracted amplitudes in a significant way for all but very thin specimens containing only light atoms.

The experimental techniques for electron scattering are largely determined by the possibility of focusing electron beams by use of strong axial magnetic fields, which act as electron lenses having focal lengths as short as 1 mm or less. Electron microscopes employing such lenses have been produced with resolutions approaching 1 Å. With such instruments, images showing individual isolated atoms of moderately high atomic number may be obtained. The resolution available is sufficient to distinguish neighbouring rows of adjacent atoms in the projected structures of thin crystals viewed in favourable orientations. It is therefore possible in many cases to obtain information on the structure of crystals and of crystal defects by direct inspection of electron micrographs.

The electromagnetic electron lenses may also be used to form electron beams of very small diameter and very high intensity. In particular, by the use of cold field-emission electron guns, it is possible to obtain a current of 10^{-10} A in an electron beam of diameter 10 Å or less with a beam divergence of less than 10^{-2} rad, *i.e.* a current density of 10^4 A cm^{-2} or more. The magnitudes of the electron scattering amplitudes then imply that detectable signals may be obtained in diffraction from assemblies of fewer than 10^2 atoms. On the other hand, electron beams may readily be collimated to better than 10^{-6} rad.

The cross sections for inelastic scattering processes are, in general, less than for the elastic scattering of electrons, but signals may be obtained by the observation of electron energy losses, or the production of secondary radiations, which allow the analysis of chemical compositions or electronic excited states for regions of the crystal 100 Å or less in diameter.

On the other hand, the transfer to the sample of large amounts of energy through inelastic scattering processes produces radiation damage which may severely limit the applicability of the imaging and diffraction techniques, especially for biological and organic materials, unless the information is gathered from large specimen volumes with low incident electron beam densities.

Structure analysis of crystals can be performed using electron diffraction in the same way as with X-ray or neutron diffraction. The mathematical expressions and the procedures are much the same. However, there are peculiarities of the electron-diffraction case which should be noted.

(1) Structure analysis based on electron diffraction is possible for thin specimens for which the conditions for kinematical scattering are approached, *e.g.* for thin mosaic single-crystal specimens, for thin polycrystalline films having a preferred orientation of very small crystallites or for very extensive, very thin single crystals of biological molecules such as membranes one or a few molecules thick.

(2) Dynamical diffraction effects are used explicitly in the determination of crystal symmetry (with no Friedel's law limitations) and for the measurement of structure amplitudes with high accuracy.

(3) For many radiation-resistant materials, the structures of crystals and of some molecules may be determined directly by imaging atom positions in projections of the crystal with a resolution of 2 Å or better. The information on atom positions is not dependent on the periodicity of the crystal and so it is equally possible to determine the structures of individual crystal defects in favourable cases.

(4) Techniques of microanalysis may be applied to the determination of the chemical composition of regions of diameter 100 Å or less using the same instrument as for diffraction, so that the chemical information may be correlated directly with morphological and structural information.

(5) Crystal-structure information may be derived from regions containing as few as 10^2 or 10^3 atoms, including very small crystals and single or multiple layers of atoms on surfaces.

The material of this section is also reviewed in the text by Spence (2003).

2.5.2.2. The interactions of electrons with matter

(1) The *elastic* scattering of electrons results from the interaction of the charged electrons with the electrostatic potential distribution, $\varphi(\mathbf{r})$, of the atoms or crystals. An incident electron of kinetic energy eW gains energy $e\varphi(\mathbf{r})$ in the potential field. Alternatively it may be stated that an incident electron wave of

[1] Questions related to this section may be addressed to Professor J. C. H. Spence (see list of contributing authors).

wavelength $\lambda = h/mv$ is diffracted by a region of variable refractive index

$$n(\mathbf{r}) = k/K_0 = \{[W + \varphi(\mathbf{r})]/W\}^{1/2} \simeq 1 + \varphi(\mathbf{r})/2W.$$

(2) The most important *inelastic* scattering processes are:

(*a*) thermal diffuse scattering, with energy losses of the order of 2×10^{-2} eV, separable from the elastic scattering only with specially devised equipment; the angular distribution of thermal diffuse scattering shows variations with $(\sin\theta)/\lambda$ which are much the same as for the X-ray case in the kinematical limit;

(*b*) bulk plasmon excitation, or the excitation of collective energy states of the conduction electrons, giving energy losses of 3 to 30 eV and an angular range of scattering of 10^{-4} to 10^{-3} rad;

(*c*) surface plasmons, or the excitation of collective energy states of the conduction electrons at discontinuities of the structure, with energy losses less than those for bulk plasmons and a similar angular range of scattering;

(*d*) interband or intraband excitation of valence-shell electrons giving energy losses in the range of 1 to 10^2 eV and an angular range of scattering of 10^{-4} to 10^{-2} rad;

(*e*) inner-shell excitations, with energy losses of 10^2 eV or more and an angular range of scattering of 10^{-3} to 10^{-2} rad, depending on the energy losses involved.

(3) In the original treatment by Bethe (1928) of the elastic scattering of electrons by crystals, the Schrödinger equation is written for electrons in the periodic potential of the crystal; *i.e.*

$$\nabla^2\psi(\mathbf{r}) + K_0^2[1 + \varphi(\mathbf{r})/W]\psi(\mathbf{r}) = 0, \tag{2.5.2.1}$$

where

$$\begin{aligned}\varphi(\mathbf{r}) &= \int V(\mathbf{u}) \exp\{-2\pi i\mathbf{u}\cdot\mathbf{r}\}\,\mathrm{d}\mathbf{u} \\ &= \sum_{\mathbf{h}} V_{\mathbf{h}} \exp\{-2\pi i\mathbf{h}\cdot\mathbf{r}\},\end{aligned} \tag{2.5.2.2}$$

$\mathbf{K}_0$ is the wavevector in zero potential (outside the crystal) (magnitude $2\pi/\lambda$) and W is the accelerating voltage. The solutions of the equation are Bloch waves of the form

$$\psi(\mathbf{r}) = \sum_{\mathbf{h}} C_{\mathbf{h}}(\mathbf{k}) \exp\{-i(\mathbf{k}_0 + 2\pi\mathbf{h})\cdot\mathbf{r}\}, \tag{2.5.2.3}$$

where $\mathbf{k}_0$ is the incident wavevector in the crystal and $\mathbf{h}$ is a reciprocal-lattice vector. Substitution of (2.5.2.2) and (2.5.2.3) in (2.5.2.1) gives the dispersion equations

$$(\kappa^2 - k_{\mathbf{h}}^2)C_{\mathbf{h}} + \sum_{\mathbf{g}}{}' V_{\mathbf{h}-\mathbf{g}}C_{\mathbf{g}} = 0. \tag{2.5.2.4}$$

Here κ is the magnitude of the wavevector in a medium of constant potential V_0 (the 'inner potential' of the crystal). The refractive index of the electron in the average crystal potential is then

$$n = \kappa/K = (1 + V_0/W)^{1/2} \simeq 1 + V_0/2W. \tag{2.5.2.5}$$

Since V_0 is positive and of the order of 10 V and W is 10^4 to 10^6 V, $n - 1$ is positive and of the order of 10^{-4}.

Solution of equation (2.5.2.4) gives the Fourier coefficients $C_{\mathbf{h}}^{(i)}$ of the Bloch waves $\psi^{(i)}(\mathbf{r})$ and application of the boundary conditions gives the amplitudes of individual Bloch waves (see Chapter 5.2).

(4) The experimentally important case of transmission of high-energy electrons through thin specimens is treated on the assumption of a plane wave incident in a direction almost perpendicular to an infinitely extended plane-parallel lamellar crystal, making use of the *small-angle scattering approximation* in which the forward-scattered wave is represented in the paraboloidal approximation to the sphere. The incident-beam direction, assumed to be almost parallel to the z axis, is unique and the z component of $\mathbf{k}$ is factored out to give

$$\nabla^2\psi + 2k\sigma\varphi\psi = \pm i2k\frac{\partial\psi}{\partial z}, \tag{2.5.2.6}$$

where $k = 2\pi/\lambda$ and $\sigma = 2\pi me\lambda/h^2$. [See Lynch & Moodie (1972), Portier & Gratias (1981), Tournarie (1962), and Chapter 5.2.]

This equation is analogous to the time-dependent Schrödinger equation with z replacing t. Retention of the $\pm$ signs on the right-hand side is consistent with both ψ and ψ^* being solutions, corresponding to propagation in opposite directions with respect to the z axis. The double-valued solution is of importance in consideration of reciprocity relationships which provide the basis for the description of some dynamical diffraction symmetries. (See Section 2.5.3.)

(5) The integral form of the wave equation, commonly used for scattering problems, is written, for electron scattering, as

$$\psi(\mathbf{r}) = \psi^{(0)}(\mathbf{r}) + (\sigma/\lambda)\int \frac{\exp\{-i\mathbf{k}|\mathbf{r} - \mathbf{r}'|\}}{|\mathbf{r} - \mathbf{r}'|}\varphi(\mathbf{r}')\psi(\mathbf{r}')\,\mathrm{d}\mathbf{r}'. \tag{2.5.2.7}$$

The wavefunction $\psi(\mathbf{r})$ within the integral is approximated by using successive terms of a Born series

$$\psi(\mathbf{r}) = \psi^{(0)}(\mathbf{r}) + \psi^{(1)}(\mathbf{r}) + \psi^{(2)}(\mathbf{r}) + \ldots. \tag{2.5.2.8}$$

The first Born approximation is obtained by putting $\psi(\mathbf{r}) = \psi^{(0)}(\mathbf{r})$ in the integral and subsequent terms $\psi^{(n)}(\mathbf{r})$ are generated by putting $\psi^{(n-1)}(\mathbf{r})$ in the integral.

For an incident plane wave, $\psi^{(0)}(\mathbf{r}) = \exp\{-i\mathbf{k}_0\cdot\mathbf{r}\}$ and for a point of observation at a large distance $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ from the scattering object ($|\mathbf{R}| \gg |\mathbf{r}'|$), the first Born approximation is generated as

$$\psi^{(1)}(\mathbf{r}) = \frac{i\sigma}{\lambda R}\exp\{-i\mathbf{k}\cdot\mathbf{R}\}\int \varphi(\mathbf{r}')\exp\{i\mathbf{q}\cdot\mathbf{r}'\}\,\mathrm{d}\mathbf{r}',$$

where $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$ or, putting $\mathbf{u} = \mathbf{q}/2\pi$ and collecting the pre-integral terms into a parameter μ,

$$\Psi(\mathbf{u}) = \mu\int \varphi(\mathbf{r})\exp\{2\pi i\mathbf{u}\cdot\mathbf{r}\}\,\mathrm{d}\mathbf{r}. \tag{2.5.2.9}$$

This is the Fourier-transform expression which is the basis for the *kinematical scattering approximation*. It is derived on the basis that all $\psi^{(n)}(\mathbf{r})$ terms for $n \neq 0$ are very much smaller than $\psi^{(0)}(\mathbf{r})$ and so is a weak scattering approximation.

In this approximation, the scattered amplitude for an atom is related to the atomic structure amplitude, $f(\mathbf{u})$, by the relationship, derived from (2.5.2.8),

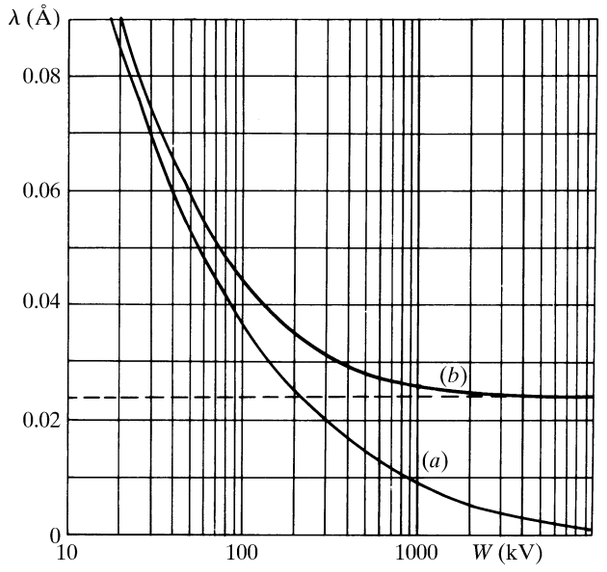


Fig. 2.5.2.1. The variation with accelerating voltage of electrons of (*a*) the wavelength, λ and (*b*) the quantity $\lambda[1 + (h^2/m_0^2c^2\lambda^2)] = \lambda_c/\beta$ which is proportional to the interaction constant σ [equation (2.5.2.14)]. The limit is the Compton wavelength λ_c (after Fujiwara, 1961).

$$\psi(\mathbf{r}) = \exp\{-i\mathbf{k}_0 \cdot \mathbf{r}\} + i\frac{\exp\{-i\mathbf{k} \cdot \mathbf{r}\}}{R\lambda}\sigma f(\mathbf{u}),$$
$$f(\mathbf{u}) = \int \varphi(\mathbf{r}) \exp\{2\pi i\mathbf{u} \cdot \mathbf{r}\}\, \mathrm{d}\mathbf{r}. \qquad (2.5.2.10)$$

For centrosymmetrical atom potential distributions, the $f(\mathbf{u})$ are real, positive and monotonically decreasing with $|\mathbf{u}|$. A measure of the extent of the validity of the first Born approximation is given by the fact that the effect of adding the higher-order terms of the Born series may be represented by replacing $f(\mathbf{u})$ in (2.5.2.10) by the complex quantities $f(\mathbf{u}) = |\mathbf{f}| \exp\{i\eta(\mathbf{u})\}$ and for single heavy atoms the phase factor η may vary from 0.2 for $|\mathbf{u}| = 0$ to 4 or 5 for large $|\mathbf{u}|$, as seen from the tables of *IT* C (2004, Section 4.3.3).

(6) Relativistic effects produce appreciable variations of the parameters used above for the range of electron energies considered. The relativistic values are

$$m = m_0(1 - v^2/c^2)^{-1/2} = m_0(1 - \beta^2)^{-1/2}, \qquad (2.5.2.11)$$
$$\lambda = h[2m_0|e|W(1 + |e|W/2m_0c^2)]^{-1/2} \qquad (2.5.2.12)$$
$$= \lambda_c(1 - \beta^2)^{1/2}/\beta, \qquad (2.5.2.13)$$

where λ_c is the Compton wavelength, $\lambda_c = h/m_0c = 0.0242$ Å, and

$$\sigma = 2\pi me\lambda/h^2 = (2\pi m_0 e/h^2)(\lambda_c/\beta)$$
$$= 2\pi/\{\lambda W[1 + (1 - \beta^2)^{1/2}]\}. \qquad (2.5.2.14)$$

Values for these quantities are listed in *IT* C (2004, Section 4.3.2). The variations of λ and σ with accelerating voltage are illustrated in Fig. 2.5.2.1. For high voltages, σ tends to a constant value, $2\pi m_0 e\lambda_c/h^2 = e/\hbar c$.

2.5.2.3. Recommended sign conventions

There are two alternative sets of signs for the functions describing wave optics. Both sets have been widely used in the literature. There is, however, a requirement for internal consistency within a particular analysis, independently of which set is adopted. Unfortunately, this requirement has not always been met and, in fact, it is only too easy at the outset of an analysis to make errors in this way. This problem might have come into prominence somewhat earlier were it not for the fact that, for centrosymmetric crystals (or indeed for centrosymmetric projections in the case of planar diffraction), only the signs used in the transmission and propagation functions can affect the results. It is not until the origin is set away from a centre of symmetry that there is a need to be consistent in every sign used.

Signs for electron diffraction have been chosen from two points of view: (1) defining as positive the sign of the exponent in the structure-factor expression and (2) defining the forward propagating free-space wavefunction with a positive exponent.

The second of these alternatives is the one which has been adopted in most solid-state and quantum-mechanical texts.

The first, or *standard crystallographic* convention, is the one which could most easily be adopted by crystallographers accustomed to retaining a positive exponent in the structure-factor equation. This also represents a consistent *International Tables* usage. It is, however, realized that both conventions will continue to be used in crystallographic computations, and that there are by now a large number of operational programs in use.

It is therefore recommended (*a*) that a particular sign usage be indicated as either *standard crystallographic* or *alternative crystallographic* to accord with Table 2.5.2.1, whenever there is a need for this to be explicit in publication, and (*b*) that either one or other of these systems be adhered to throughout an analysis in a self-consistent way, even in those cases where, as indicated above, some of the signs appear to have no effect on one particular conclusion.

2.5.2.4. Scattering of electrons by crystals; approximations

The forward-scattering approximation to the many-beam dynamical diffraction theory outlined in Chapter 5.2 provides the basis for the calculation of diffraction intensities and electron-microscope image contrast for thin crystals. [See Cowley (1995), Chapter 5.2 and *IT* C (2004) Sections 4.3.6 and 4.3.8.] On the other hand, there are various approximations which provide relatively simple analytical expressions, are useful for the determination of diffraction geometry, and allow estimates to be made of the relative intensities in diffraction patterns and electron micrographs in favourable cases.

(*a*) *The kinematical approximation*, derived in Section 2.5.2.2 from the first Born approximation, is analagous to the corresponding approximation of X-ray diffraction. It assumes that the scattering amplitudes are directly proportional to the three-dimensional Fourier transform of the potential distribution, $\varphi(\mathbf{r})$.

$$V(\mathbf{u}) = \int \varphi(\mathbf{r}) \exp\{2\pi i\mathbf{u} \cdot \mathbf{r}\}\, \mathrm{d}\mathbf{r}, \qquad (2.5.2.15)$$

so that the potential distribution $\varphi(\mathbf{r})$ takes the place of the charge-density distribution, $\rho(\mathbf{r})$, relevant for X-ray scattering.

The validity of the kinematical approximation as a basis for structure analysis is severely limited. For light-atom materials, such as organic compounds, it has been shown by Jap & Glaeser (1980) that the thickness for which the approximation gives reasonable accuracy for zone-axis patterns from single crystals is of the order of 100 Å for 100 keV electrons and increases, approximately as σ^{-1}, for higher energies. The thickness limits quoted for polycrystalline samples, having crystallite dimensions smaller than the sample thickness, are usually greater (Vainshtein, 1956). For heavy-atom materials the approximation is more limited since it may fail significantly for single heavy atoms.

(*b*) *The phase-object approximation* (POA), or high-voltage limit, is derived from the general many-beam dynamical

Table 2.5.2.1. *Standard crystallographic and alternative crystallographic sign conventions for electron diffraction*

	Standard	Alternative
Free-space wave	$\exp[-i(\mathbf{k}\cdot\mathbf{r}-\omega t)]$	$\exp[+i(\mathbf{k}\cdot\mathbf{r}-\omega t)]$
Fourier transforming from real space to reciprocal space	$\int \psi(\mathbf{r})\exp[+2\pi i(\mathbf{u}\cdot\mathbf{r})]\,\mathrm{d}\mathbf{r}$	$\int \psi(\mathbf{r})\exp[-2\pi i(\mathbf{u}\cdot\mathbf{r})]\,\mathrm{d}\mathbf{r}$
Fourier transforming from reciprocal space to real space	$\psi(\mathbf{r}) = \int \Psi(\mathbf{u})\exp[-2\pi i(\mathbf{u}\cdot\mathbf{r})]\,\mathrm{d}\mathbf{u}$	$\int \Psi(\mathbf{u})\exp[+2\pi i(\mathbf{u}\cdot\mathbf{r})]\,\mathrm{d}\mathbf{u}$
Structure factors	$V(\mathbf{h}) = (1/\Omega)\sum_j f_j(\mathbf{h})\exp(+2\pi i\mathbf{h}\cdot\mathbf{r}_j)$	$(1/\Omega)\sum_j f_j(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{r}_j)$
Transmission function (real space)	$\exp[-i\sigma\varphi(x,y)\Delta z]$	$\exp[+i\sigma\varphi(x,y)\Delta z]$
Phenomenological absorption	$\sigma\varphi(\mathbf{r}) - i\mu(\mathbf{r})$	$\sigma\varphi(\mathbf{r}) + i\mu(\mathbf{r})$
Propagation function $P(h)$ (reciprocal space) within the crystal	$\exp(-2\pi i\zeta_{\mathbf{h}}\Delta z)$	$\exp(+2\pi i\zeta_{\mathbf{h}}\Delta z)$
Iteration (reciprocal space)	$\Psi_{n+1}(\mathbf{h}) = [\Psi_n(\mathbf{h})\cdot P(\mathbf{h})] * Q(\mathbf{h})$	
Unitarity test (for no absorption)	$T(\mathbf{h}) = Q(\mathbf{h}) * Q^*(-\mathbf{h}) = \delta(\mathbf{h})$	
Propagation to the image plane-wave aberration function, where $\chi(U) = \pi\lambda\Delta f U^2 + \frac{1}{2}\pi C_s\lambda^3 U^4$, $U^2 = u^2 + v^2$ and Δf is positive for overfocus	$\exp[i\chi(U)]$	$\exp[-i\chi(U)]$

σ = electron interaction constant $= 2\pi me\lambda/h^2$; m = (relativistic) electron mass; λ = electron wavelength; e = (magnitude of) electron charge; h = Planck's constant; $k = 2\pi/\lambda$; Ω = volume of the unit cell; $\mathbf{u}$ = continuous reciprocal-space vector, components u, v; $\mathbf{h}$ = discrete reciprocal-space coordinate; $\varphi(x, y)$ = crystal potential averaged along beam direction (positive); Δz = slice thickness; $\mu(\mathbf{r})$ = absorption potential [positive; typically $\leq 0.1\sigma\varphi(\mathbf{r})$]; Δf = defocus (defined as negative for underfocus); C_s = spherical aberration coefficient; $\zeta_{\mathbf{h}}$ = excitation error relative to the incident-beam direction and defined as negative when the point h lies outside the Ewald sphere; $f_j(\mathbf{h})$ = atomic scattering factor for electrons, f_e, related to the atomic scattering factor for X-rays, f_X, by the Mott formula $f_e = (e/\pi U^2)(Z - f_X)$. $Q(\mathbf{h})$ = Fourier transform of periodic slice transmission function.

diffraction expression, equation (5.2.13.1), Chapter 5.2, by assuming the Ewald sphere curvature to approach zero. Then the scattering by a thin sample can be expressed by multiplying the incoming wave amplitude by the transmission function

$$q(xy) = \exp\{-i\sigma\varphi(xy)\}, \tag{2.5.2.16}$$

where $\varphi(xy) = \int \varphi(\mathbf{r})\,\mathrm{d}z$ is the projection of the potential distribution of the sample in the z direction, the direction of the incident beam. The diffraction-pattern amplitudes are then given by two-dimensional Fourier transform of (2.5.2.16).

This approximation is of particular value in relation to the electron microscopy of thin crystals. The thickness for its validity for 100 keV electrons is within the range 10 to 50 Å, depending on the accuracy and spatial resolution involved, and increases with accelerating voltage approximately as $\lambda^{-1/2}$. In computational work, it provides the starting point for the multislice method of dynamical diffraction calculations (*IT* C, 2004, Section 4.3.6.1).

(*c*) *The two-beam approximation* for dynamical diffraction of electrons assumes that only two beams, the incident beam and one diffracted beam (or two Bloch waves, each with two component amplitudes), exist in the crystal. This approximation has been adapted, notably by Hirsch *et al.* (1965), for use in the electron microscopy of inorganic materials.

It forms a convenient basis for the study of defects in crystals having small unit cells (metals, semiconductors *etc.*) and provides good preliminary estimates for the determination of crystal thicknesses and structure amplitudes for orientations well removed from principal axes, and for electron energies up to 200–500 keV, but it has decreasing validity, even for favourable cases, for higher energies. It has been used in the past as an 'extinction correction' for powder-pattern intensities (Vainshtein, 1956).

(*d*) *The Bethe second approximation*, proposed by Bethe (1928) as a means for correcting the two-beam approximation for the effects of weakly excited beams, replaces the Fourier coefficients of potential by the 'Bethe potentials'

$$U_{\mathbf{h}} = V_{\mathbf{h}} - 2k_0\sigma\sum_{\mathbf{g}}\frac{V_{\mathbf{g}}\cdot V_{\mathbf{h}-\mathbf{g}}}{\kappa^2 - k_{\mathbf{g}}^2}. \tag{2.5.2.17}$$

Use of these potentials has been shown to account well for the deviations of powder-pattern intensities from the predictions of two-beam theory (Horstmann & Meyer, 1965) and to predict accurately the extinctions of Kikuchi lines at particular accelerating voltages due to relativistic effects (Watanabe *et al.*, 1968), but they give incorrect results for the small-thickness limit.

2.5.2.5. Kinematical diffraction formulae

(1) *Comparison with X-ray diffraction.* The relations of real-space and reciprocal-space functions are analogous to those for X-ray diffraction [see equations (2.5.2.2), (2.5.2.10) and (2.5.2.15)]. For diffraction by crystals

$$\varphi(\mathbf{r}) = \sum_{\mathbf{h}} V_{\mathbf{h}}\exp\{-2\pi i\mathbf{h}\cdot\mathbf{r}\},$$

$$V_{\mathbf{h}} = \int \varphi(\mathbf{r})\exp\{2\varphi i\mathbf{h}\cdot\mathbf{r}\}\,\mathrm{d}\mathbf{r} \tag{2.5.2.18}$$

$$= \frac{1}{\Omega}\sum_i f_i(\mathbf{h})\exp\{2\pi i\mathbf{h}\cdot\mathbf{r}_i\}, \tag{2.5.2.19}$$

where the integral of (2.5.2.18) and the summation of (2.5.2.19) are taken over one unit cell of volume (see Dawson *et al.*, 1974).

Important differences from the X-ray case arise because

(*a*) the wavelength is relatively small so that the Ewald-sphere curvature is small in the reciprocal-space region of appreciable scattering amplitude;

(*b*) the dimensions of the single-crystal regions giving appreciable scattering amplitudes are small so that the 'shape transform' regions of scattering power around the reciprocal-lattice points are relatively large;

(*c*) the spread of wavelengths is small (10^{-5} or less, with no white-radiation background) and the degree of collimation is better (10^{-4} to 10^{-6}) than for conventional X-ray sources.

As a consequence of these factors, single-crystal diffraction patterns may show many simultaneous reflections, representing almost-planar sections of reciprocal space, and may show fine structure or intensity variations reflecting the crystal dimensions and shape.

(2) Kinematical diffraction-pattern intensities are calculated in a manner analogous to that for X-rays except that

(*a*) no polarization factor is included because of the small-angle scattering conditions;

(*b*) integration over regions of scattering power around reciprocal-lattice points cannot be assumed unless appropriate experimental conditions are ensured.

For a thin, flat, lamellar crystal of thickness H, the observed intensity is

$$I_{\mathbf{h}}/I_{\mathbf{0}} = |\sigma(V_{\mathbf{h}}/\Omega)(\sin \pi\zeta_{\mathbf{h}} H)/(\pi\zeta_{\mathbf{h}})|^2, \tag{2.5.2.20}$$

where $\zeta_{\mathbf{h}}$ is the excitation error for the $\mathbf{h}$ reflection and Ω is the unit-cell volume.

For a single-crystal diffraction pattern obtained by rotating a crystal or from a uniformly bent crystal or for a mosaic crystal with a uniform distribution of orientations, the intensity is

$$I_{\mathbf{h}} = I_{\mathbf{0}} \frac{\sigma^2 |V_{\mathbf{h}}|^2 V_{\mathrm{c}} d_{\mathbf{h}}}{4\pi^2 \Omega^2}, \tag{2.5.2.21}$$

where V_{c} is the crystal volume and $d_{\mathbf{h}}$ is the lattice-plane spacing. For a polycrystalline sample of randomly oriented small crystals, the intensity per unit length of the diffraction ring is

$$I_{\mathbf{h}} = I_{\mathbf{0}} \frac{\sigma^2 |V_{\mathbf{h}}|^2 V_{\mathrm{c}} d_{\mathbf{h}}^2 M_{\mathbf{h}}}{8\pi^2 \Omega^2 L\lambda}, \tag{2.5.2.22}$$

where $M_{\mathbf{h}}$ is the multiplicity factor for the $\mathbf{h}$ reflection and L is the camera length, or the distance from the specimen to the detector plane. The special cases of 'oblique texture' patterns from powder patterns having preferred orientations are treated in *IT* C (2004, Section 4.3.5).

(3) *Two-beam dynamical diffraction formulae: complex potentials including absorption.* In the two-beam dynamical diffraction approximation, the intensities of the directly transmitted and diffracted beams for transmission through a crystal of thickness H, in the absence of absorption, are

$$I_{\mathbf{0}} = (1 + w^2)^{-1} \left[w^2 + \cos^2 \left\{ \frac{\pi H (1 + w^2)^{1/2}}{\xi_{\mathbf{h}}} \right\} \right] \tag{2.5.2.23}$$

$$I_{\mathbf{h}} = (1 + w^2)^{-1} \sin^2 \left\{ \frac{\pi H (1 + w^2)^{1/2}}{\xi_{\mathbf{h}}} \right\}, \tag{2.5.2.24}$$

where $\xi_{\mathbf{h}}$ is the extinction distance, $\xi_{\mathbf{h}} = (2\sigma|V_{\mathbf{h}}|)^{-1}$, and

$$w = \xi_{\mathbf{h}} \zeta_{\mathbf{h}} = \Delta\theta/(2\sigma|V_{\mathbf{h}}|d_{\mathbf{h}}), \tag{2.5.2.25}$$

where $\Delta\theta$ is the deviation from the Bragg angle.

For the case that $\zeta_{\mathbf{h}} = 0$, with the incident beam at the Bragg angle, this reduces to the simple *Pendellösung* expression

$$I_{\mathbf{h}} = 1 - I_{\mathbf{0}} = \sin^2\{2\pi\sigma|V_{\mathbf{h}}|H\}. \tag{2.5.2.26}$$

The effects on the elastic Bragg scattering amplitudes of the inelastic or diffuse scattering may be introduced by adding an out-of-phase component to the structure amplitudes, so that for a centrosymmetric crystal, $V_{\mathbf{h}}$ becomes complex by addition of an imaginary component. Alternatively, an absorption function $\mu(\mathbf{r})$, having Fourier coefficients $\mu_{\mathbf{h}}$, may be postulated so that $\sigma V_{\mathbf{h}}$ is replaced by $\sigma V_{\mathbf{h}} + i\mu_{\mathbf{h}}$. The $\mu_{\mathbf{h}}$ are known as phenomenological absorption coefficients and their validity in many-beam diffraction has been demonstrated by, for example, Rez (1978).

The magnitudes $\mu_{\mathbf{h}}$ depend on the nature of the experiment and the extent to which the various inelastically or diffusely scattered electrons are included in the measurements being made. If measurements are made of purely elastic scattering intensities for Bragg reflections or of image intensity variations due to the interaction of the sharp Bragg reflections only, the main contributions to the absorption coefficients are as follows (Radi, 1970):

(*a*) from plasmon and single-electron excitations, $\mu_{\mathbf{0}}$ is of the order of 0.1 $V_{\mathbf{0}}$ and $\mu_{\mathbf{h}}$, for $\mathbf{h} \neq 0$, is negligibly small;

(*b*) from thermal diffuse scattering; $\mu_{\mathbf{h}}$ is of the order of 0.1 $V_{\mathbf{h}}$ and decreasing more slowly than $V_{\mathbf{h}}$ with scattering angle.

Including absorption effects in (2.5.2.26) for the case $\zeta_{\mathbf{h}} = 0$ gives

$$\begin{aligned} I_{\mathbf{0}} &= \tfrac{1}{2}\exp\{-\mu_{\mathbf{0}} H\}[\cosh \mu_{\mathbf{h}} H + \cos(2\pi\sigma V_{\mathbf{h}} H)], \\ I_{\mathbf{h}} &= \tfrac{1}{2}\exp\{-\mu_{\mathbf{0}} H\}[\cosh \mu_{\mathbf{h}} H - \cos(2\pi\sigma V_{\mathbf{h}} H)]. \end{aligned} \tag{2.5.2.27}$$

The Borrmann effect is not very pronounced for electrons because $\mu_{\mathbf{h}} \ll \mu_{\mathbf{0}}$, but can be important for the imaging of defects in thick crystals (Hirsch *et al.*, 1965; Hashimoto *et al.*, 1961).

Attempts to obtain analytical solutions for the dynamical diffraction equations for more than two beams have met with few successes. There are some situations of high symmetry, with incident beams in exact zone-axis orientations, for which the many-beam solution can closely approach equivalent two- or three-beam behaviour (Fukuhara, 1966). Explicit solutions for the three-beam case, which displays some aspects of many-beam character, have been obtained (Gjønnes & Høier, 1971; Hurley & Moodie, 1980).

2.5.2.6. Imaging with electrons

Electron optics. Electrons may be focused by use of axially symmetric magnetic fields produced by electromagnetic lenses. The focal length of such a lens used as a projector lens (focal points outside the lens field) is given by

$$f_{\mathrm{p}}^{-1} = \frac{e}{8mW_{\mathrm{r}}} \int_{-\infty}^{\infty} H_z^2(z)\,\mathrm{d}z, \tag{2.5.2.28}$$

where W_{r} is the relativistically corrected accelerating voltage and H_z is the z component of the magnetic field. An expression in terms of experimental constants was given by Liebman (1955) as

$$\frac{1}{f} = \frac{A_0 (NI)^2}{W_{\mathrm{r}}(S + D)}, \tag{2.5.2.29}$$

where A_0 is a constant, NI is the number of ampere turns of the lens winding, S is the length of the gap between the magnet pole pieces and D is the bore of the pole pieces.

Lenses of this type have irreducible aberrations, the most important of which for the paraxial conditions of electron microscopy is the third-order spherical aberration, coefficient C_{s}, giving a variation of focal length of $C_{\mathrm{s}}\alpha^2$ for a beam at an angle α to the axis. Chromatic aberration, coefficient C_{c}, gives a spread of focal lengths

$$\Delta f = C_{\mathrm{c}} \left(\frac{\Delta W_0}{W_0} + 2\frac{\Delta I}{I} \right) \tag{2.5.2.30}$$

for variations ΔW_0 and ΔI of the accelerating voltage and lens currents, respectively.

The objective lens of an electron microscope is the critical lens for the determination of image resolution and contrast. The action of this lens in a conventional transmission electron microscope (TEM) is described by use of the Abbe theory for coherent incident illumination transmitted through the object to produce a wavefunction $\psi_0(xy)$ (see Fig. 2.5.2.2).

The amplitude distribution in the back focal plane of the objective lens is written

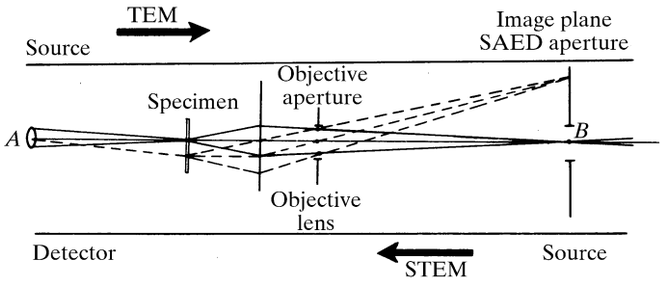


Fig. 2.5.2.2. Diagram representing the critical components of a conventional transmission electron microscope (TEM) and a scanning transmission electron microscope (STEM). For the TEM, electrons from a source A illuminate the specimen and the objective lens forms an image of the transmitted electrons on the image plane, B. For the STEM, a source at B is imaged by the objective lens to form a small probe on the specimen and some part of the transmitted beam is collected by a detector at A.

$$\Psi_0(u, v) \cdot T(u, v), \tag{2.5.2.31}$$

where $\Psi_0(u, v)$ is the Fourier transform of $\psi_0(x, y)$ and $T(u, v)$ is the transfer function of the lens, consisting of an aperture function

$$A(u, v) = \begin{cases} 1 & \text{for } (u^2 + v^2)^{1/2} \leq A \\ 0 & \text{elsewhere} \end{cases} \tag{2.5.2.32}$$

and a phase function $\exp\{i\chi(u, v)\}$ where the phase perturbation $\chi(uv)$ due to lens defocus Δf and aberrations is usually approximated as

$$\chi(uv) = \pi \cdot \Delta f \cdot \lambda(u^2 + v^2) + \frac{\pi}{2} C_s \lambda^3 (u^2 + v^2)^2, \tag{2.5.2.33}$$

and u, v are the reciprocal-space variables related to the scattering angles φ_x, φ_y by

$$u = (\sin \varphi_x)/\lambda,$$
$$v = (\sin \varphi_y)/\lambda.$$

The image amplitude distribution, referred to the object coordinates, is given by Fourier transform of (2.5.2.31) as

$$\psi(xy) = \psi_0(xy) * t(xy), \tag{2.5.2.34}$$

where $t(xy)$, given by Fourier transform of $T(u, v)$, is the spread function. The image intensity is then

$$I(xy) = |\psi(xy)|^2 = |\psi_0(xy) * t(xy)|^2. \tag{2.5.2.35}$$

In practice the coherent imaging theory provides a good approximation but limitations of the coherence of the illumination have appreciable effects under high-resolution imaging conditions.

The variation of focal lengths according to (2.5.2.30) is described by a function $G(\Delta f)$. Illumination from a finite incoherent source gives a distribution of incident-beam angles $H(u_1, v_1)$. Then the image intensity is found by integrating incoherently over Δf and u_1, v_1:

$$I(xy) = \int\int G(\Delta f) \cdot H(u_1 v_1) \times |\mathscr{F}\{\Psi_0(u - u_1, v - v_1) \cdot T_{\Delta f}(u, v)\}|^2 \, \mathrm{d}(\Delta f) \cdot \mathrm{d}u_1 \, \mathrm{d}v_1, \tag{2.5.2.36}$$

where $\mathscr{F}$ denotes the Fourier-transform operation.

In the scanning transmission electron microscope (STEM), the objective lens focuses a small bright source of electrons on the object and directly transmitted or scattered electrons are detected to form an image as the incident beam is scanned over the object (see Fig. 2.5.2.2). Ideally the image amplitude can be related to that of the conventional transmission electron microscope by use of the 'reciprocity relationship' which refers to point sources and detectors for scalar radiation in scalar fields with elastic scattering processes only. It may be stated: 'The amplitude at a point B due to a point source at A is identical to that which would be produced at A for the identical source placed at B'.

For an axial point source, the amplitude distribution produced by the objective lens on the specimen is

$$\mathscr{F}[T(u, v)] = t(xy). \tag{2.5.2.37}$$

If this is translated by the scan to X, Y, the transmitted wave is

$$\psi_0(xy) = q(xy) \cdot t(x - X, y - Y). \tag{2.5.2.38}$$

The amplitude on the plane of observation following the specimen is then

$$\Psi(uv) = Q(u, v) * \{T(uv) \exp[2\pi i(uX + vY)]\}, \tag{2.5.2.39}$$

and the image signal produced by a detector having a sensitivity function $H(u, v)$ is

$$I(X, Y) = \int H(u, v) |Q(u, v) * T(u, v) \times \exp\{2\pi i(uX + vY)\}|^2 \, \mathrm{d}u \, \mathrm{d}v. \tag{2.5.2.40}$$

If $H(u, v)$ represents a small detector, approximated by a delta function, this becomes

$$I(x, y) = |q(xy) * t(xy)|^2, \tag{2.5.2.41}$$

which is identical to the result (2.5.2.35) for a plane incident wave in the conventional transmission electron microscope.

2.5.2.7. Imaging of very thin and weakly scattering objects

(*a*) *The weak-phase-object approximation.* For sufficiently thin objects, the effect of the object on the incident-beam amplitude may be represented by the transmission function (2.5.2.16) given by the phase-object approximation. If the fluctuations, $\varphi(xy) - \bar{\varphi}$, about the mean value of the projected potential are sufficiently small so that $\sigma[\varphi(xy) - \bar{\varphi}] \ll 1$, it is possible to use the *weak-phase-object approximation* (WPOA)

$$q(xy) = \exp\{-i\sigma\varphi(xy)\} = 1 - i\sigma\varphi(xy), \tag{2.5.2.42}$$

where $\varphi(xy)$ is referred to the average value, $\bar{\varphi}$. The assumption that only first-order terms in $\sigma\varphi(xy)$ need be considered is the equivalent of a single-scattering, or kinematical, approximation applied to the two-dimensional function, the projected potential

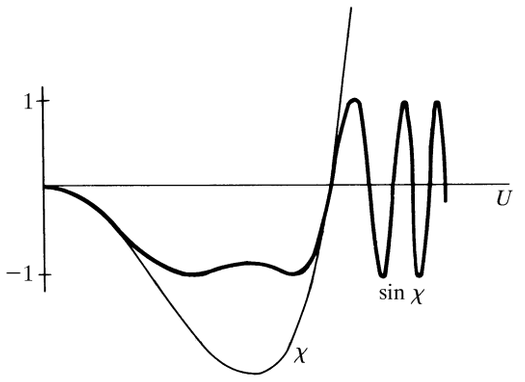


Fig. 2.5.2.3. The functions $\chi(U)$, the phase factor for the transfer function of a lens given by equation (2.5.2.33), and $\sin\chi(U)$ for the Scherzer optimum defocus condition, relevant for weak phase objects, for which the minimum value of $\chi(U)$ is $-2\pi/3$.

of (2.5.2.16). From (2.5.2.42), the image intensity (2.5.2.35) becomes

$$I(xy) = 1 + 2\sigma\varphi(xy) * s(xy), \tag{2.5.2.43}$$

where the spread function $s(xy)$ is the Fourier transform of the imaginary part of $T(uv)$, namely $A(uv)\sin\chi(uv)$.

The optimum imaging condition is then found, following Scherzer (1949), by specifying that the defocus should be such that $|\sin\chi|$ is close to unity for as large a range of $U = (u^2 + v^2)^{1/2}$ as possible. This is so for a negative defocus such that $\chi(uv)$ decreases to a minimum of about $-2\pi/3$ before increasing to zero and higher as a result of the fourth-order term of (2.5.2.33) (see Fig. 2.5.2.3). This optimum, 'Scherzer defocus' value is given by

$$\frac{\mathrm{d}\chi}{\mathrm{d}u} = 0 \quad \text{for} \quad \chi = -2\pi/3$$

or

$$\Delta f = -\left(\tfrac{4}{3} C_\mathrm{s}\lambda\right)^{1/2}. \tag{2.5.2.44}$$

The resolution limit is then taken as corresponding to the value of $U = 1.51 C_\mathrm{s}^{-1/4}\lambda^{-3/4}$ when $\sin\chi$ becomes zero, before it begins to oscillate rapidly with U. The resolution limit is then

$$\Delta x = 0.66 C_\mathrm{s}^{1/4}\lambda^{3/4}. \tag{2.5.2.45}$$

For example, for $C_\mathrm{s} = 1$ mm and $\lambda = 2.51 \times 10^{-2}$ Å (200 keV), $\Delta x = 2.34$ Å.

Within the limits of the WPOA, the image intensity can be written simply for a number of other imaging modes in terms of the Fourier transforms $c(\mathbf{r})$ and $s(\mathbf{r})$ of the real and imaginary parts of the objective-lens transfer function $T(\mathbf{u}) = A(\mathbf{u})\exp\{i\chi(\mathbf{u})\}$, where $\mathbf{r}$ and $\mathbf{u}$ are two-dimensional vectors in real and reciprocal space, respectively.

For dark-field TEM images, obtained by introducing a central stop to block out the central beam in the diffraction pattern in the back-focal plane of the objective lens,

$$I(\mathbf{r}) = [\sigma\varphi(\mathbf{r}) * c(\mathbf{r})]^2 + [\sigma\varphi(\mathbf{r}) * s(\mathbf{r})]^2. \tag{2.5.2.46}$$

Here, as in (2.5.2.42), $\varphi(\mathbf{r})$ should be taken to imply the difference from the mean potential value, $\varphi(\mathbf{r}) - \bar{\varphi}$.

For bright-field STEM imaging with a very small detector placed axially in the central beam of the diffraction pattern (2.5.2.39) on the detector plane, the intensity, from (2.5.2.41), is given by (2.5.2.43).

For a finite axially symmetric detector, described by $D(\mathbf{u})$, the image intensity is

$$I(\mathbf{r}) = 1 + 2\sigma\varphi(\mathbf{r}) * \{s(\mathbf{r})[d(\mathbf{r}) * c(\mathbf{r})] - c(\mathbf{r})[d(\mathbf{r}) * s(\mathbf{r})]\}, \tag{2.5.2.47}$$

where $d(\mathbf{r})$ is the Fourier transform of $D(\mathbf{u})$ (Cowley & Au, 1978).

For STEM with an annular dark-field detector which collects all electrons scattered outside the central spot of the diffraction pattern in the detector plane, it can be shown that, to a good approximation (valid except near the resolution limit)

$$I(\mathbf{r}) = \sigma^2\varphi^2(\mathbf{r}) * [c^2(\mathbf{r}) + s^2(\mathbf{r})]. \tag{2.5.2.48}$$

Since $c^2(\mathbf{r}) + s^2(\mathbf{r}) = |t(\mathbf{r})|^2$ is the intensity distribution of the electron probe incident on the specimen, (2.5.2.48) is equivalent to the incoherent imaging of the function $\sigma^2\varphi^2(\mathbf{r})$.

Within the range of validity of the WPOA or, in general, whenever the zero beam of the diffraction pattern is very much stronger than any diffracted beam, the general expression (2.5.2.36) for the modifications of image intensities due to limited coherence may be conveniently approximated. The effect of integrating over the variables $\Delta f, u_1, v_1$, may be represented by multiplying the transfer function $T(u, v)$ by so-called 'envelope functions' which involve the Fourier transforms of the functions $G(\Delta f)$ and $H(u_1, v_1)$.

For example, if $G(\Delta f)$ is approximated by a Gaussian of width ε (at e^{-1} of the maximum) centred at Δf_0 and $H(u_1 v_1)$ is a circular aperture function

$$H(u_1 v_1) = \begin{cases} 1 & \text{if } u_1, v_1 < b \\ 0 & \text{otherwise,} \end{cases}$$

the transfer function $T_0(uv)$ for coherent radiation is multiplied by

$$\exp\{-\pi^2\lambda^2\varepsilon^2(u^2 + v^2)^2/4\} \cdot J_1(\pi B\eta)/(\pi B\eta)$$

where

$$\eta = f_0\lambda(u + v) + C_\mathrm{s}\lambda^3(u^3 + v^3) + \pi i\varepsilon^2\lambda^2(u^3 + u^2 v + uv^2 + v^3)/2. \tag{2.5.2.49}$$

(*b*) *The projected charge-density approximation.* For very thin specimens composed of moderately heavy atoms, the WPOA is inadequate. Within the region of validity of the phase-object approximation (POA), more complicated relations analagous to (2.5.2.43) to (2.5.2.47) may be written. A simpler expression may be obtained by use of the two-dimensional form of Poisson's equation, relating the projected potential distribution $\varphi(xy)$ to the projected charge-density distribution $\rho(xy)$. This is the PCDA (projected charge-density approximation) (Cowley & Moodie, 1960),

$$I(xy) = 1 + 2\Delta f \cdot \lambda\sigma\rho(xy). \tag{2.5.2.50}$$

This is valid for sufficiently small values of the defocus Δf, provided that the effects of the spherical aberration may be neglected, *i.e.* for image resolutions not too close to the Scherzer

resolution limit (Lynch *et al.*, 1975). The function $\rho(xy)$ includes contributions from both the positive atomic nuclei and the negative electron clouds. For underfocus (Δf negative), single atoms give dark spots in the image. The contrast reverses with defocus.

2.5.2.8. Crystal structure imaging

(*a*) *Introduction.* It follows from (2.5.2.43) and (2.5.2.42) that, within the severe limitations of validity of the WPOA or the PCDA, images of very thin crystals, viewed with the incident beam parallel to a principal axis, will show dark spots at the positions of rows of atoms parallel to the incident beam. Provided that the resolution limit is less than the projected distances between atom rows (1–3 Å), the projection of the crystal structure may be seen directly.

In practice, theoretical and experimental results suggest that images may give a direct, although nonlinear, representation of the projected potential or charge-density distribution for thicknesses much greater than the thicknesses for validity of these approximations, *e.g.* for thicknesses which may be 50 to 100 Å for 100 keV electrons for 3 Å resolutions and which increase for comparable resolutions at higher voltage but decrease with improved resolutions.

The use of high-resolution imaging as a means for determining the structures of crystals and for investigating the form of the defects in crystals in terms of the arrangement of the atoms has become a widely used and important branch of crystallography with applications in many areas of solid-state science. It must be emphasized, however, that image intensities are strongly dependent on the crystal thickness and orientation and also on the instrumental parameters (defocus, aberrations, alignment *etc.*). It is only when all of these parameters are correctly adjusted to lie within strictly defined limits that interpretation of images in terms of atom positions can be attempted [see *IT* C (2004, Section 4.3.8)].

Reliable interpretations of high-resolution images of crystals ('crystal structure images') may be made, under even the most favourable circumstances, only by the comparison of observed image intensities with intensities calculated by use of an adequate approximation to many-beam dynamical diffraction theory [see *IT* C (2004, Section 4.3.6)]. Most calculations for moderate or large unit cells are currently made by the multislice method based on formulation of the dynamical diffraction theory due to Cowley & Moodie (1957). For smaller unit cells, the matrix method based on the Bethe (1928) formulation is also frequently used (Hirsch *et al.*, 1965).

(*b*) *Fourier images.* For periodic objects in general, and crystals in particular, the amplitudes of the diffracted waves in the back focal plane are given from (2.5.2.31) by

$$\Psi_0(\mathbf{h}) \cdot T(\mathbf{h}). \qquad (2.5.2.51)$$

For rectangular unit cells of the projected unit cell, the vector $\mathbf{h}$ has components h/a and k/b. Then the set of amplitudes (2.5.2.34), and hence the image intensities, will be identical for two different sets of defocus and spherical aberration values Δf_1, C_{s1} and Δf_2, C_{s2} if, for an integer N,

$$\chi_1(h) = \chi_2(h) = 2N\pi;$$

i.e.

$$\pi\lambda\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right)(\Delta f_1 - \Delta f_2) + \frac{1}{2}\pi\lambda^3\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right)^2 (C_{s1} - C_{s2}) = 2\pi N.$$

This relationship is satisfied for all h, k if a^2/b^2 is an integer and

$$\Delta f_1 - \Delta f_2 = 2na^2/\lambda$$

and

$$C_{s1} - C_{s2} = 4ma^4/\lambda^3, \qquad (2.5.2.52)$$

where m, n are integers (Kuwabara, 1978). The relationship for Δf is an expression of the Fourier image phenomenon, namely that for a plane-wave incidence, the intensity distribution for the image of a periodic object repeats periodically with defocus (Cowley & Moodie, 1960). Hence it is often necessary to define the defocus value by observation of a nonperiodic component of the specimen such as a crystal edge (Spence *et al.*, 1977).

For the special case $a^2 = b^2$, the image intensity is also reproduced exactly for

$$\Delta f_1 - \Delta f_2 = (2n+1)a^2/\lambda, \qquad (2.5.2.53)$$

except that in this case the image is translated by a distance $a/2$ parallel to each of the axes.

2.5.2.9. Image resolution

(1) *Definition and measurement.* The 'resolution' of an electron microscope or, more correctly, the 'least resolvable distance', is usually defined by reference to the transfer function for the coherent imaging of a weak phase object under the Scherzer optimum defocus condition (2.5.2.44). The resolution figure is taken as the inverse of the U value for which $\sin\chi(U)$ first crosses the axis and is given, as in (2.5.2.45), by

$$\Delta x = 0.66C_s^{1/4}\lambda^{3/4}. \qquad (2.5.2.45)$$

It is assumed that an objective aperture is used to eliminate the contribution to the image for U values greater than the first zero crossing, since for these contributions the relative phases are distorted by the rapid oscillations of $\sin\chi(U)$ and the corresponding detail of the image is not directly interpretable in terms of the projection of the potential distribution of the object.

The resolution of the microscope in practice may be limited by the incoherent factors which have the effect of multiplying the WPOA transfer function by envelope functions as in (2.5.2.49).

The resolution, as defined above, and the effects of the envelope functions may be determined by Fourier transform of the image of a suitable thin, weakly scattering amorphous specimen. The Fourier-transform operation may be carried out by use of an optical diffractometer. A more satisfactory practice is to digitize the image directly by use of a two-dimensional detector system in the microscope or from a photographic recording, and perform the Fourier transform numerically.

For the optical diffractometer method, the intensity distribution obtained is given from (2.5.2.43) as a radially symmetric function of U,

$$\begin{aligned} I(U) &= |\mathscr{F} I(xy)|^2 \\ &= \delta(U) + 4\sigma^2|\Phi(u)|^2 \cdot \sin^2\chi(U) \cdot E^2(U), \end{aligned} \qquad (2.5.2.54)$$

where $E(U)$ is the product of the envelope functions.

In deriving (2.5.2.54) it has been assumed that:

(*a*) the WPOA applies;

(*b*) the optical transmission function of the photographic record is linearly related to the image intensity, $I(xy)$;

(*c*) the diffraction intensity $|\Phi(U)|^2$ is a radially symmetric, smoothly varying function such as is normally produced by a sufficiently large area of the image of an amorphous material;

(*d*) there is no astigmatism present and no drift of the specimen; either of these factors would remove the radial symmetry.

From the form of (2.5.2.54) and a preknowledge of $|\Phi(U)|^2$, the zero crossings of $\sin\chi$ and the form of $E(U)$ may be deduced. Analysis of a through-focus series of images provides more complete and reliable information.

(2) Detail on a scale much smaller than the resolution of the electron microscope, as defined above, is commonly seen in electron micrographs, especially for crystalline samples. For example, lattice fringes, having the periodicity of the crystal lattice planes, with spacings as small as 0.6 Å in one direction, have been observed using a microscope having a resolution of about 2.5 Å (Matsuda *et al.*, 1978), and two-dimensionally periodic images showing detail on the scale of 0.5 to 1 Å have been observed with a similar microscope (Hashimoto *et al.*, 1977).

Such observations are possible because

(*a*) for periodic objects the diffraction amplitude $\Psi_0(uv)$ in (2.5.2.31) is a set of delta functions which may be multiplied by the corresponding values of the transfer function that will allow strong interference effects between the diffracted beams and the zero beam, or between different diffracted beams;

(*b*) the envelope functions for the WPOA, arising from incoherent imaging effects, do not apply for strongly scattering crystals; the more general expression (2.5.2.36) provides that the incoherent imaging factors will have much less effect on the interference of some sets of diffracted beams.

The observation of finely spaced lattice fringes provides a measure of some important factors affecting the microscope performance, such as the presence of mechanical vibrations, electrical interference or thermal drift of the specimen. A measure of the fineness of the detail observable in this type of image may therefore be taken as a measure of 'instrumental resolution'.

2.5.2.10. Electron diffraction in electron microscopes

Currently most electron-diffraction patterns are obtained in conjunction with images, in electron microscopes of one form or another, as follows.

(*a*) Selected-area electron-diffraction (SAED) patterns are obtained by using intermediate and projector lenses to form an image of the diffraction pattern in the back-focal plane of the objective lens (Fig. 2.5.2.2). The area of the specimen from which the diffraction pattern is obtained is defined by inserting an aperture in the image plane of the objective lens. For parallel illumination of the specimen, sharp diffraction spots are produced by perfect crystals.

A limitation to the area of the specimen from which the diffraction pattern can be obtained is imposed by the spherical aberration of the objective lens. For a diffracted beam scattered through an angle α, the spread of positions in the object for which the diffracted beam passes through a small axial aperture in the image plane is $C_s\alpha^3$, *e.g.* for $C_s = 1$ mm, $\alpha = 5 \times 10^{-2}$ rad (10,0,0 reflection from gold for 100 keV electrons), $C_s\alpha^3 = 1250$ Å, so that a selected-area diameter of less than about 2000 Å is not feasible. For higher voltages, the minimum selected-area diameter decreases with λ^2 if the usual assumption is made that C_s increases for higher-voltage microscopes so that $C_s\lambda$ is a constant.

(*b*) Convergent-beam electron-diffraction (CBED) patterns are obtained when an incident convergent beam is focused on the specimen, as in an STEM instrument or an STEM attachment for a conventional TEM instrument.

For a large, effectively incoherent source, such as a conventional hot-filament electron gun, the intensities are added for each incident-beam direction. The resulting CBED pattern has an intensity distribution

$$I(uv) = \int |\Psi_{u_1v_1}(uv)|^2 \, \mathrm{d}u_1 \, \mathrm{d}v_1, \qquad (2.5.2.55)$$

where $\Psi_{u_1v_1}(uv)$ is the Fourier transform of the exit wave at the specimen for an incident-beam direction u_1, v_1.

(*c*) Coherent illumination from a small bright source such as a field emission gun may be focused on the specimen to give an electron probe having an intensity distribution $|t(xy)|^2$ and a diameter equal to the STEM dark-field image resolution [equation (2.5.2.47)] of a few Å. The intensity distribution of the resulting microdiffraction pattern is then

$$|\Psi(uv)|^2 = |\Psi_0(uv) * T(uv)|^2, \qquad (2.5.2.56)$$

where $\Psi_0(uv)$ is the Fourier transform of the exit wave at the specimen. Interference occurs between waves scattered from the various incident-beam directions. The diffraction pattern is thus an in-line hologram as envisaged by Gabor (1949).

(*d*) Diffraction patterns may be obtained by using an optical diffractometer (or computer) to produce the Fourier transform squared of a small selected region of a recorded image. The optical diffraction-pattern intensity obtained under the ideal conditions specified under equation (2.5.2.54) is given, in the case of weak phase objects, by

$$I(uv) = \delta(uv) + 4\sigma^2|\Phi(uv)|^2 \cdot \sin^2\chi(uv) \cdot E^2(uv) \qquad (2.5.2.57)$$

or, more generally, by

$$I(uv) = c\delta(uv) + |\Psi(uv) \cdot T(uv) * \Psi^*(uv) \cdot T^*(uv)|^2,$$

where $\Psi(uv)$ is the Fourier transform of the wavefunction at the exit face of the specimen and c is a constant depending on the characteristics of the photographic recording medium.

2.5.3. Point-group and space-group determination by convergent-beam electron diffraction

By M. Tanaka

2.5.3.1. Introduction

Because the cross section for electron scattering is at least a thousand times greater than that for X-rays, and because multiple Bragg scattering preserves information on symmetry (such as the absence of inversion symmetry), electron diffraction is exquisitely sensitive to symmetry. The additional ability of modern electron-optical lenses to focus an electron probe down to nanometre dimensions, and so allow the study of nanocrystals too small for analysis by X-rays, has meant that the method of convergent-beam diffraction described here has now become the preferred method of symmetry determination for very small crystals, domains, twinned structures, quasicrystals, incommensurate structures and other imperfectly crystalline materials.

Convergent-beam electron diffraction (CBED) originated with the experiments of Kossel & Möllenstedt (1938). However, modern crystallographic investigations by CBED began with the studies performed by Goodman & Lehmpfuhl (1965) in a modified transmission electron microscope. They obtained CBED patterns by converging a conical electron beam with an angle of more than 10^{-3} rad on an ~30 nm diameter specimen area, which had uniform thickness and no bending. Instead of the

usual diffraction spots, diffraction discs (in Laue or transmission geometry) were produced. The diffraction intensity within a disc shows a specific symmetry, which enables one to determine the point groups and space groups of microcrystals. Unlike X-ray diffraction, the method is extremely sensitive to the presence or absence of inversion symmetry.

The method corresponding to CBED in the field of light optics is the conoscope method. Using a conoscope, we can identify whether a crystal is isotropic, uniaxial or biaxial, and determine the optic axis and the sign of birefringence of a crystal. When CBED, a conoscope method using an electron beam, is utilized, more basic properties of a crystal – the crystal point group and space group – can be determined.

Point- and space-group determinations are routinely also carried out by X-ray diffraction. This method, to which kinematical diffraction is applicable, cannot determine whether a crystal is polar or nonpolar unless anomalous absorption is utilized. As a result, the X-ray diffraction method can only identify 11 Laue groups among 32 point groups. CBED, based fully upon dynamical diffraction, can distinguish polar crystals from nonpolar crystals using only a nanometre-sized crystal, thus allowing the unique identification of all the point groups by inspecting the symmetries appearing in CBED discs.

As pointed out above, an unambiguous experimental determination of crystal symmetry, in the case of X-ray diffraction, is usually not possible because of the apparent centrosymmetry of the diffraction pattern, even for noncentrosymmetric crystals. However, methods based on structure-factor and X-ray intensity statistics remain useful for the resolution of space-group ambiguities, and are routinely applied to structure determinations from X-ray data. These methods are described in Chapter 2.1 of this volume.

In the field of materials science, correct space-group determination by CBED is often requested prior to X-ray or neutron structure refinement, in particular in the case of Rietveld refinements based on powder diffraction data.

CBED can determine not only the point and space groups of crystals but also crystal structure parameters – lattice parameters, atom positions, Debye–Waller factors and low-order structure factors. The lattice parameters can be determined from submicron regions of thin crystals by using higher-order Laue zone (HOLZ) reflections with an accuracy of 1×10^{-4}. Cherns *et al.* (1988) were the first to perform strain analysis of artificial multilayer materials using the large-angle technique (LACBED) (Tanaka *et al.*, 1980). Since then, many strain measurements at interfaces of various multilayer materials have been successfully conducted. In recent years, strain analysis has been conducted using automatic analysis programs, which take account of dynamical diffraction effects (Krämer *et al.*, 2000). We refer to the book of Morniroli (2002), which carries many helpful figures, clear photographs and a comprehensive list of papers on this topic.

Vincent *et al.* (1984*a*,*b*) first applied the CBED method to the determination of the atom positions of AuGeAs. They analysed the intensities of HOLZ reflections by applying a quasi-kinematical approximation. Tanaka & Tsuda (1990, 1991) and Tsuda & Tanaka (1995) refined the structural parameters of $SrTiO_3$ by applying the dynamical theory of electron diffraction. The method was extended to the refinements of CdS, $LaCrO_3$ and hexagonal $BaTiO_3$ (Tsuda & Tanaka, 1999; Tsuda *et al.*, 2002; Ogata *et al.*, 2004). Rossouw *et al.* (1996) measured the order parameters of TiAl through a Bloch-wave analysis of HOLZ reflections in a CBED pattern. Midgley *et al.* (1996) refined two positional parameters of $AuSn_4$ from the diffraction data obtained with a small convergence angle using multislice calculations.

Low-order structure factors were first determined by Goodman & Lehmpfuhl (1967) for MgO. After much work on low-order structure-factor determination, Zuo & Spence determined the 200 and 400 structure factors of MgO in a very modern way, by fitting energy-filtered patterns and many-beam dynamical calculations using a least-squares procedure. For the low-order structure-factor determinations, the excellent comprehensive review of Spence (1993) should be referred to. Saunders *et al.* (1995) succeeded in obtaining the deformation charge density of Si using the low-order crystal structure factors determined by CBED. For the reliable determination of the low-order X-ray crystal structure factors or the charge density of a crystal, accurate determination of the Debye–Waller factors is indispensable. Zuo *et al.* (1999) determined the bond-charge distribution in cuprite. Simultaneous determination of the Debye–Waller factors and the low-order structure factors using HOLZ and zeroth-order Laue zone (ZOLZ) reflections was performed to determine the deformation charge density of $LaCrO_3$ accurately (Tsuda *et al.*, 2002).

CBED can also be applied to the determination of lattice defects, dislocations (Cherns & Preston, 1986), stacking faults (Tanaka, 1986) and twins (Tanaka, 1986). Since this topic is beyond the scope of the present chapter, readers are referred to pages 156 to 205 of the book by Tanaka *et al.* (1994).

We also mention the book by Spence & Zuo (1992), which deals with the whole topic of CBED, including the basic theory and a wealth of literature.

2.5.3.2. Point-group determination

When an electron beam traverses a thin slab of crystal parallel to a zone axis, one can easily imagine that symmetries parallel to the zone axis should appear in the resulting CBED pattern. It is, however, more difficult to imagine what symmetries appear due to symmetries perpendicular to the incident beam. Goodman (1975) pioneered the clarification of CBED symmetries for the twofold rotation axis and mirror plane perpendicular to the incident beam, and the symmetry of an inversion centre, with the help of the reciprocity theorem of scattering theory. Tinnappel (1975) solved many CBED symmetries at various crystal settings with respect to the incident beam using a group-theoretical treatment. Buxton *et al.* (1976) also derived these results from first principles, and generalized them to produce a systematic method for the determination of the crystal point group. Tanaka, Saito & Sekii (1983) developed a method to determine the point group using simultaneously excited many-beam patterns. The point-group-determination method given by Buxton *et al.* (1976) is described with the aid of the description by Tanaka, Saito & Sekii (1983) in the following.

2.5.3.2.1. Symmetry elements of a specimen and diffraction groups

Since CBED uses the Laue geometry, Buxton *et al.* (1976) assumed a perfectly crystalline specimen in the form of a parallel-sided slab which is infinite in two dimensions. The symmetry elements of the *specimen* (as distinct from those of an infinite crystal) form 'diffraction groups', which are isomorphic to the point groups of the diperiodic plane figures and Shubnikov groups of coloured plane figures. The diffraction groups of a specimen are determined from the symmetries of CBED patterns taken at various orientations of the specimen. The *crystal* point-group of the specimen is identified by referring to Fig. 2.5.3.4, which gives the relation between diffraction groups and crystal point groups.

A specimen that is parallel-sided and is infinitely extended in the x and y directions has ten symmetry elements. The symmetry elements consist of six two-dimensional symmetry elements and four three-dimensional ones. The operation of the former elements transforms an arbitrary coordinate (x, y, z) into (x', y', z), with z remaining the same. The operation of the latter transforms a coordinate (x, y, z) into (x', y', z'), where $z' \neq z$. A vertical mirror plane m and one-, two-, three-, four- and sixfold rotation axes that are parallel to the surface normal **z** are the two-

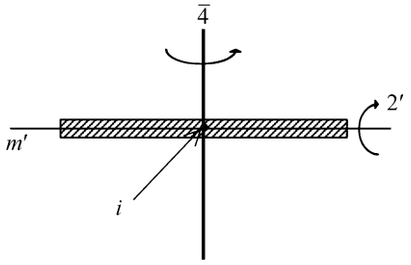


Fig. 2.5.3.1. Four symmetry elements m', i, $2'$ and $\bar{4}$ of an infinitely extended parallel-sided specimen.

Table 2.5.3.1. *Two- and three-dimensional symmetry elements of an infinitely extended parallel-sided specimen*

Symbols in parentheses show CBED symmetries appearing in dark-field patterns.

Two-dimensional symmetry elements	Three-dimensional symmetry elements
1	m' (1_R)
2	i (2_R)
3	$2'$ (m_2, m_R)
4	$\bar{4}$ (4_R)
5	
6	
m	

dimensional symmetry elements. A horizontal mirror plane m', an inversion centre i, a horizontal twofold rotation axis $2'$ and a fourfold rotary inversion $\bar{4}$ are the three-dimensional symmetry elements, and are shown in Fig. 2.5.3.1. The fourfold rotary inversion was not recognized as a symmetry element until the point groups of the diperiodic plane figures were considered (Buxton *et al*., 1976). Table 2.5.3.1 lists these symmetry elements, where the symbols in parentheses express symmetries of CBED patterns expected from three-dimensional symmetry elements.

The diffraction groups are constructed by combining these symmetry elements (Table 2.5.3.2). Two-dimensional symmetry elements and their combinations are given in the top row of the table. The third symmetry m in parentheses is introduced automatically when the first two symmetry elements are combined. Three-dimensional symmetry elements are given in the first column. The equations given below the table indicate that no additional three-dimensional symmetry elements can appear by combination of two symmetry elements in the first column. As a result, 31 diffraction groups are produced by combining the elements in the first column with those in the top row. Diffraction groups in square brackets have already appeared ealier in the table. In the fourth row, three columns have two diffraction groups, which are produced when symmetry elements are combined at different orientations. In the last row, five columns are empty because a fourfold rotary inversion cannot coexist with threefold and sixfold rotation axes. In the last column, the number of independent diffraction groups in each row is given, the sum of the numbers being 31.

2.5.3.2.2. Identification of three-dimensional symmetry elements

It is difficult to imagine the symmetries in CBED patterns generated by the three-dimensional symmetry elements of the sample. The reason is that if a three-dimensional symmetry element is applied to a specimen, it turns it upside down, which is impractical in most experiments. The reciprocity theorem of scattering theory (Pogany & Turner, 1968) enables us to clarify the symmetries of CBED patterns expected from these three-dimensional symmetry elements. A graphical method for obtaining CBED symmetries due to sample symmetry elements is described in the papers of Goodman (1975), Buxton *et al*. (1976) and Tanaka (1989). The CBED symmetries of the three-dimensional symmetries do not appear in the zone-axis patterns, but do in a diffraction disc set at the Bragg condition, each of which we call a dark-field pattern (DP). The CBED symmetries obtained are illustrated in Fig. 2.5.3.2. A horizontal twofold rotation axis $2'$, a horizontal mirror plane m', an inversion centre i and a fourfold rotary inversion $\bar{4}$ produce symmetries m_R (m_2), 1_R, 2_R and 4_R in DPs, respectively.

Next we explain the symbols of the CBED symmetries. (1) Operation m_R is shown in the left-hand part of Fig. 2.5.3.2(*a*), which implies successive operations of (*a*) a mirror m with respect to a twofold rotation axis, transforming an open circle beam (○) in reflection G into a beam (+) in reflection G' and (*b*) rotation R of this beam by π about the centre point of disc G' (or the exact Bragg position of reflection G'), resulting in position ○ in reflection G'. The combination of the two operations is written as m_R. When the twofold rotation axis is parallel to the diffraction vector **G**, two beams (○) in the left-hand part of the figure become one reflection G, and a mirror symmetry, whose mirror line is perpendicular to vector **G** and passes through the centre of disc G, appears between the two beams (the right-hand side figure of Fig. 2.5.3.2*a*). The mirror symmetry is labelled m_2 after the twofold rotation axis. (2) Operation 1_R (Fig. 2.5.3.2*b*) for a horizontal mirror plane is a combination of a rotation by 2π of a beam (○) about a zone axis O (symbol 1), which is equivalent to no rotation, and a rotation by π of the beam about the exact Bragg position or the centre of disc G. (3) Operation 2_R is a rotation by π of a beam (○) in reflection G about a zone axis (symbol 2), which transforms the beam into a beam (+) in reflection $-G$, followed by a rotation by π of the beam (+) about the centre of disc $-G$, resulting in the beam (○) in disc $-G$ (Fig. 2.5.3.2*c*). The symmetry is called translational symmetry after Goodman (1975) because the pattern of disc $+G$ coincides with that of disc $-G$ by a translation. It is emphasized that an inversion centre is identified by the test of translational symmetry about a pair of $\pm G$ dark-field patterns – if one disc can be translated into coincidence with the other, an inversion centre exists. We call the pair ±DP. (4) Operation 4_R (Fig. 2.5.3.2*d*) can be understood in a similar manner. It is noted that regular letters are symmetries about a zone axis, while subscripts R represent symmetries about the exact Bragg position. We call a pattern that contains an exact Bragg position (if possible at the disc centre) a dark-field pattern. As far as CBED symmetries are concerned, we

Table 2.5.3.2. *Symmetry elements of an infinitely extended parallel-sided specimen and diffraction groups*

	1	2	3	4	6	m	$2m(m)$	$3m$	$4m(m)$	$6m(m)$	
1	1	2	3	4	6	m	$2m(m)$	$3m$	$4m(m)$	$6m(m)$	10
(m') 1_R	1_R	21_R	31_R	41_R	61_R	$m1_R$	$2m(m)1_R$	$3m1_R$	$4m(m)1_R$	$6m(m)1_R$	10
(i) 2_R	2_R	[21_R]	6_R	[41_R]	[61_R]	$2_Rm(m_R)$	[$2m(m)1_R$]	$6_Rm(m_R)$	[$4m(m)1_R$]	[$6m(m)1_R$]	4
						[$2_Rm(m_R)$]	[$2m(m)1_R$]	[$3m1_R$]			
($2'$) m_R	m_R	$2m_R(m_R)$	$3m_R$	$4m_R(m_R)$	$6m_R(m_R)$	[$m1_R$]	[$4_Rm(m)m_R$]	[$6_Rm(m_R)$]	[$4m(m)1_R$]	[$6_Rm(m_R)$]	5
($\bar{4}$) 4_R		4_R		[41_R]		$4_Rm(m_R)$	[$4_Rm(m_R)$]		[$4m(m)1_R$]		2

$1_R \times 2_R = 2$, $2_R \times 2_R = 1$, $m_R \times 2_R = m$, $4_R \times 2_R = 4$, $1_R \times m_R = m \times m_R$, $1_R \times 4_R = 4 \times 1_R$, $m_R \times 4_R = m \times 4_R$.

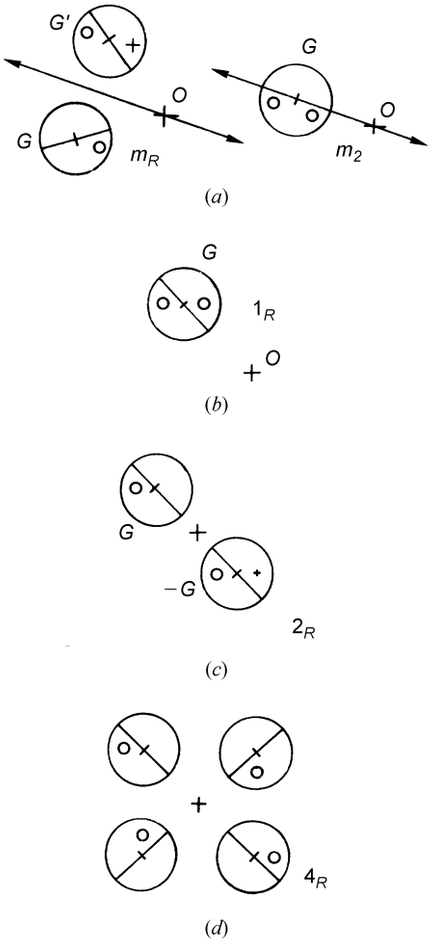


Fig. 2.5.3.2. Illustration of symmetries appearing in dark-field patterns (DPs). (*a*) m_R and m_2; (*b*) 1_R; (*c*) 2_R; (*d*) 4_R, originating from $2'$, m', i and $\bar{4}$, respectively.

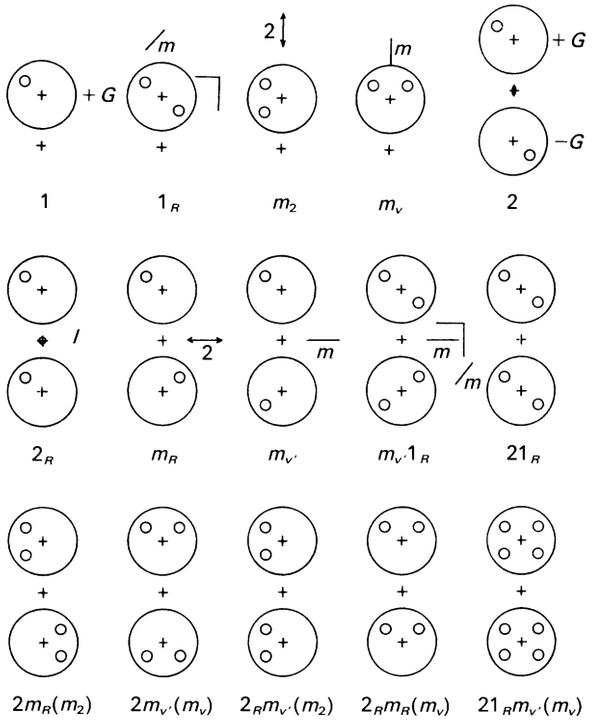


Fig. 2.5.3.3. Illustration of symmetries appearing in dark-field patterns (DPs) and a pair of dark-field patterns (±DP) for the combinations of symmetry elements.

do not use the term dark-field pattern if a disc does not contain the exact Bragg position.

The four three-dimensional symmetry elements are found to produce different symmetries in the DPs. These facts imply that these symmetry elements can be identified unambiguously from the symmetries of CBED patterns.

2.5.3.2.3. Identification of two-dimensional symmetry elements

Two-dimensional symmetry elements that belong to a zone axis exhibit their symmetries in CBED patterns or zone-axis patterns (ZAPs) directly, even if dynamical diffraction takes place. A ZAP contains a bright-field pattern (BP) and a whole pattern (WP). The BP is the pattern appearing in the bright-field disc [the central or 'direct' (000) beam]. The WP is composed of the BP and the pattern formed by the surrounding diffraction discs, which are not exactly excited. The two-dimensional symmetry elements m, 1, 2, 3, 4 and 6 yield symmetry m_v and one-, two-, three-, four- and sixfold rotation symmetries, respectively, in WPs, where the suffix v for m_v is assigned to distinguish it from mirror symmetry m_2 caused by a horizontal twofold rotation axis.

It should be noted that a BP shows not only the zone-axis symmetry but also three-dimensional symmetries, indicating that the BP can have a higher symmetry than the symmetry of the corresponding WP. The symmetries of the BP due to three-dimensional symmetry elements are obtained by moving the DPs to the zone axis. As a result, the three-dimensional symmetry elements $m', i, 2'$ and $\bar{4}$ produce, respectively, symmetries 1_R, 1, m_2 and 4 in the BP, instead of 1_R, 2_R, m_2 and 4_R in the DPs (Fig. 2.5.3.2). We mention that the BP cannot distinguish whether a specimen crystal has an inversion centre or not, because an inversion centre forms the lowest symmetry 1 in the BP.

In conclusion, all the two-dimensional symmetry elements can be identified from the WP symmetries.

2.5.3.2.4. Diffraction-group determination

All the symmetry elements of the diffraction groups can be identified from the symmetries of a WP and DPs. But it is practical and convenient to use just the four patterns WP, BP, DP and ±DP to determine the diffraction group. The symmetries appearing in these four patterns are given for the 31 diffraction groups in Table 2.5.3.3 (Tanaka, Saito & Sekii, 1983), which is a detailed version of Table 2 of Buxton *et al.* (1976). All the possible symmetries of the DP and ±DP appearing at different crystal orientations are given in the present table. When a BP has a higher symmetry than the corresponding WP, the symmetry elements that produce the BP are given in parentheses in column II except only for the case of 4_R. When two types of vertical mirror planes exist, these are distinguished by symbols m_v and $m_{v'}$. Each of the two or three symmetries given in columns IV and V for many diffraction groups appears in a DP or ±DP in different directions.

It is emphasized again that no two diffraction groups exhibit the same combination of BP, WP, DP and ±DP, which implies that the diffraction groups are uniquely determined from an inspection of these pattern symmetries. Fig. 2.5.3.3 illustrates the symmetries of the DP and ±DP appearing in Table 2.5.3.3, which greatly eases the cumbersome task of determining the symmetries. The first four patterns illustrate the symmetries appearing in a single DP and the others treat those in ±DPs. The pattern symmetries are written beneath the figures. The other symbols are the symmetries of a specimen. The crosses outside the diffraction discs designate the zone axis. The crosses inside the diffraction discs indicate the exact Bragg position.

When the four patterns appearing in three photographs are taken and examined using Table 2.5.3.3 with the aid of Fig. 2.5.3.3, one diffraction group can be selected unambiguously. It is,

Table 2.5.3.3. *Symmetries of different patterns for diffraction and projection diffraction groups*

(II) Bright-field patterns (BPs); (III) whole patterns (WPs); (IV) dark-field patterns (DPs); and (V) ±dark-field patterns (±DPs) for diffraction groups (I) and projection diffraction groups (VI).

I	II	III	IV	V	VI
1	1	1	1	1	1_R
1_R	2 (1_R)	1	$2 = 1_R$	1	
2	2	2	1	2	21_R
2_R	1	1	1	2_R	
21_R	2	2	2	21_R	
m_R	m (m_2)	1	1	1	$m1_R$
				m_R	
			m_2	1	
m	m_v	m_v	1	1	
				m_v	
			m_v	1	
$m1_R$	$2mm$ [$m_v + m_2 + (1_R)$]	m_v	2	1	
				m_v1_R	
			$2m_vm_2$	1	
$2m_Rm_R$	$2mm$ ($2 + m_2$)	2	1	2	$2mm1_R$
			m_2	$2m_R(m_2)$	
$2mm$	$2m_vm_{v'}$	$2m_vm_{v'}$	1	2	
			m_v	$2m_{v'}(m_v)$	
2_Rmm_R	m_v	m_v	1	2_R	
			m_2	$2_Rm_{v'}(m_2)$	
			m_v	$2_Rm_R(m_v)$	
$2mm1_R$	$2m_vm_{v'}$	$2m_vm_{v'}$	2	21_R	
			$2m_vm_2$	$21_Rm_{v'}(m_v)$	
4	4	4	1	2	41_R
4_R	4	2	1	2	
41_R	4	4	2	21_R	
$4m_Rm_R$	$4mm$ ($4 + m_2$)	4	1	2	$4mm1_R$
			m_2	$2m_R(m_2)$	
$4mm$	$4m_vm_{v'}$	$4m_vm_{v'}$	1	2	
			m_v	$2m_{v'}(m_v)$	
4_Rmm_R	$4mm$ ($2m_vm_{v'} + m_2$)	$2m_vm_{v'}$	1	2	
			m_2	$2m_R(m_2)$	
			m_v	$2m_{v'}(m_v)$	
$4mm1_R$	$4m_vm_{v'}$	$4m_vm_{v'}$	2	21_R	
			$2m_vm_2$	$21_Rm_{v'}(m_v)$	
3	3	3	1	1	31_R
31_R	6 ($3 + 1_R$)	3	2	1	
$3m_R$	$3m$ ($3 + m_2$)	3	1	1	$3m1_R$
				m_R	
			m_2	1	
$3m$	$3m_v$	$3m_v$	1	1	
				m_v	
			m_v	1	
$3m1_R$	$6mm$ [$3m_v + m_2 + (1_R)$]	$3m_v$	2	1	
				m_v1_R	
			$2m_vm_2$	1	
6	6	6	1	2	61_R
6_R	3	3	1	2_R	
61_R	6	6	2	21_R	

Table 2.5.3.3 (*cont.*)

I	II	III	IV	V	VI
$6m_Rm_R$	$6mm$ ($6 + m_2$)	6	1	2	$6mm1_R$
			m_2	$2m_R(m_2)$	
$6mm$	$6m_vm_{v'}$	$6m_vm_{v'}$	1	2	
			m_v	$2m_{v'}(m_v)$	
6_Rmm_R	$3m_v$	$3m_v$	1	2_R	
			m_2	$2_Rm_{v'}(m_2)$	
			m_v	$2_Rm_R(m_v)$	
$6mm1_R$	$6m_vm_{v'}$	$6m_vm_{v'}$	2	21_R	
			$2m_vm_2$	$21_Rm_{v'}(m_v)$	

however, noted that many diffraction groups are determined from a WP and BP pair without using a DP or ±DP (or from one photograph) or from a set of a WP, a BP and a DP without using a ±DP (or from two photographs).

2.5.3.2.5. Point-group determination

Fig. 2.5.3.4 provides the relationship between the 31 diffraction groups for slabs and the 32 point groups for infinite crystals given by Buxton *et al.* (1976). When a diffraction group is determined, possible point groups are selected by consulting this figure. Each of the 11 high-symmetry diffraction groups corresponds to only one crystal point group. In this case, the point group is uniquely determined from the diffraction group. When more than one point group falls under a diffraction group, a different diffraction group has to be obtained for another zone axis. A point group is identified by finding a common point group among the point groups obtained for different zone axes. It is clear from the figure that high-symmetry zones should be used for quick determination of point groups because low-symmetry zone axes exhibit only a small portion of crystal symmetries in the CBED patterns. Furthermore, it should be noted that CBED cannot observe crystal symmetries oblique to an incident beam or horizontal three-, four- or sixfold rotation axes. The diffraction groups to be expected for different zone axes are given for all the point groups in Table 2.5.3.4 (Buxton *et al.*, 1976). The table is useful for finding a suitable zone axis to distinguish candidate point groups expected in advance.

We shall explain the point-group determination procedure using an Si crystal. Fig. 2.5.3.5(*a*) shows a [111] ZAP of the Si specimen. The BP has threefold rotation symmetry and mirror symmetry or symmetry $3m_v$, which are caused by the threefold rotation axis along the [111] direction and a vertical mirror plane. The WP has the same symmetry. Figs. 2.5.3.5(*b*) and (*c*) are $2\bar{2}0$ and $\bar{2}20$ DPs, respectively. Both show symmetry m_2 perpendicular to the reflection vector. This symmetry is caused by a twofold rotation axis parallel to the specimen surface. One DP coincides with the other upon translation. This translational or 2_R symmetry indicates the existence of an inversion centre. By consulting Table 2.5.3.3, the diffraction group giving rise to these pattern symmetries is found to be 6_Rmm_R. Fig. 2.5.3.4 shows that there are two point groups $\bar{3}m$ and $m\bar{3}m$ causing diffraction group 6_Rmm_R. Fig. 2.5.3.6 shows another ZAP, which shows symmetry $4mm$ in the BP and the WP. The point group which has fourfold rotation symmetry is not $\bar{3}m$ but $m\bar{3}m$. The point group of Si is thus determined to be $m\bar{3}m$.

2.5.3.2.6. Projection diffraction groups

HOLZ reflections appear as excess HOLZ rings far outside the ZOLZ reflection discs and as deficit lines in the ZOLZ discs. By ignoring these weak diffraction effects with components along the beam direction, we may obtain information about the symmetry of the sample as projected along the beam direction. Thus when HOLZ reflections are weak and no deficit HOLZ

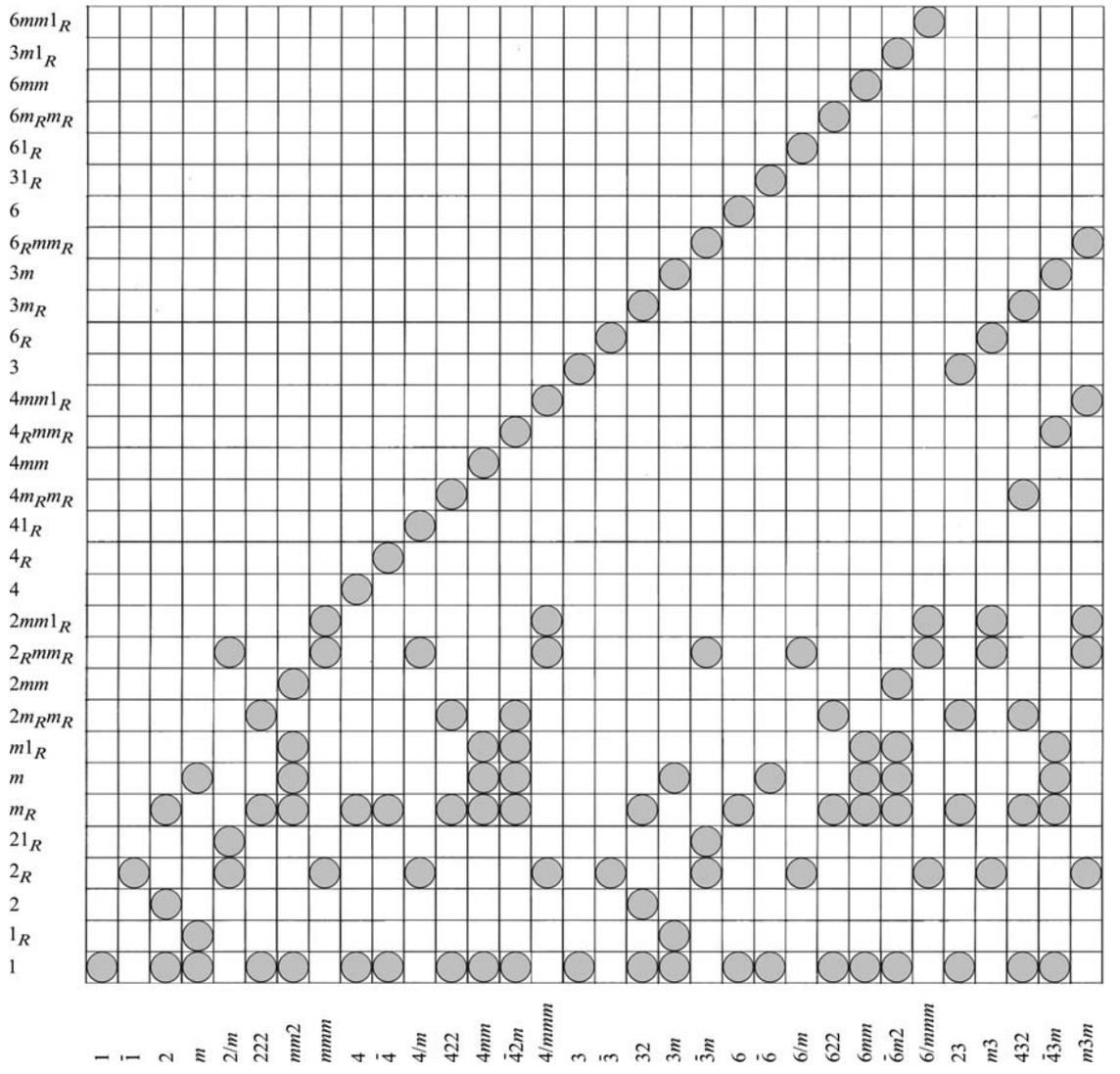


Fig. 2.5.3.4. Relation between diffraction groups and crystal point groups (after Buxton *et al.*, 1976).

lines are seen in the ZOLZ discs, the symmetry elements found from the CBED patterns are only those of the specimen projected along the zone axis. The projection of the specimen along the zone axis causes horizontal mirror symmetry m', the corresponding CBED symmetry being 1_R. When symmetry 1_R is added to the 31 diffraction groups, ten projection diffraction groups having symmetry symbol 1_R are derived as shown in column VI of Table 2.5.3.3. If only ZOLZ reflections are observed in CBED patterns, a projection diffraction group instead of a diffraction group is obtained, where only the pattern symmetries given in the rows of the diffraction groups having symmetry symbol 1_R in Table 2.5.3.3 should be consulted. Two projection diffraction groups obtained from two different zone axes are the minimum needed to determine a crystal point group, because it is constructed by the three-dimensional combination of symmetry elements. It should be noted that if a diffraction group is determined carelessly from CBED patterns with no HOLZ lines, the wrong crystal point group is obtained.

2.5.3.2.7. Symmetrical many-beam method

In the sections above, the point-group determination method established by Buxton *et al.* (1976) was described, where two- and three-dimensional symmetry elements were determined, respectively, from ZAPs and DPs.

The Laue circle is defined as the intersection of the Ewald sphere with the ZOLZ, and all reflections on this circle are simultaneously at the Bragg condition. If many such DPs are recorded (all simultaneously at the Bragg condition), many three-dimensional symmetry elements can be identified from one photograph. Using a group-theoretical method, Tinnappel (1975) studied the symmetries appearing in simultaneously excited DPs for various combinations of crystal symmetry elements. Based upon his treatment, Tanaka, Saito & Sekii (1983) developed a method for determining diffraction groups using simultaneously excited symmetrical hexagonal six-beam, square four-beam and rectangular four-beam CBED patterns. All the CBED symmetries appearing in the symmetrical many-beam (SMB) patterns were derived by the graphical method used in the paper of Buxton *et al.* (1976). From an experimental viewpoint, it is advantageous that symmetry elements can be identified from one photograph. It was found that twenty diffraction groups can be identified from one SMB pattern, whereas ten diffraction groups can be determined by Buxton *et al.*'s method. An experimental comparison between the two methods was performed by Howe *et al.* (1986).

SMB patterns are easily obtained by tilting a specimen crystal or the incident beam from a zone axis into an orientation to excite low-order reflections simultaneously. Fig. 2.5.3.7 illustrates the symmetries of the SMB patterns for all the diffraction groups except for the five groups 1, 1_R, 2, 2_R and 21_R. For these groups, the two-beam method for exciting one reflection is satisfactory because many-beam excitation gives no more information than the two-beam case. In the six-beam and square four-beam cases,

Table 2.5.3.4. *Diffraction groups expected at various crystal orientations for 32 point groups*

This table is adapted from Buxton *et al.* (1976).

	Zone-axis symmetries					
Point group	⟨111⟩	⟨100⟩	⟨110⟩	⟨*uv*0⟩	⟨*uuw*⟩	[*uvw*]
$m3m$	$6_R mm_R$	$4mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	2_R
$\bar{4}3m$	$3m$	$4_R mm_R$	$m1_R$	m_R	m	1
432	$3m_R$	$4m_R m_R$	$2m_R m_R$	m_R	m_R	1

	Zone-axis symmetries			
Point group	⟨111⟩	⟨100⟩	⟨*uv*0⟩	[*uvw*]
$m3$	6_R	$2mm1_R$	$2_R mm_R$	2_R
23	3	$2m_R m_R$	m_R	1

	Zone-axis symmetries						
Point group	[0001]	⟨11$\bar{2}$0⟩	⟨1$\bar{1}$00⟩	[*uv*.0]	[*uu*.*w*]	[*u*$\bar{u}$.*w*]	[*uv*.*w*]
$6/mmm$	$6mm1_R$	$2mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm$	$2_R mm_R$	2_R
$\bar{6}m2$	$3m1_R$	$m1_R$	$2mm$	m	m_R	m	1
$6mm$	$6mm$	$m1_R$	$m1_R$	m_R	m	m	1
622	$6m_R m_R$	$2m_R m_R$	$2m_R m_R$	m_R	m_R	m_R	1

	Zone-axis symmetries		
Point group	[0001]	[*uv*.0]	[*uv*.*w*]
$6/m$	61_R	$2_R mm_R$	2_R
$\bar{6}$	31_R	m	1
6	6	m_R	1

	Zone-axis symmetries			
Point group	[0001]	⟨11$\bar{2}$0⟩	[*u*$\bar{u}$.*w*]	[*uv*.*w*]
$\bar{3}m$	$6_R mm_R$	21_R	$2_R mm_R$	2_R
$3m$	$3m$	1_R	m	1
32	$3m_R$	2	m_R	1

	Zone-axis symmetries	
Point group	[0001]	[*uv*.*w*]
$\bar{3}$	6_R	2_R
3	3	1

	Zone-axis symmetries						
Point group	[001]	⟨100⟩	⟨110⟩	[*u*0*w*]	[*uv*0]	[*uuw*]	[*uvw*]
$4/mmm$	$4mm1_R$	$2mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	$2_R mm_R$	2_R
$\bar{4}2m$	$4_R mm_R$	$2m_R m_R$	$m1_R$	m_R	m_R	m	1
$4mm$	$4mm$	$m1_R$	$m1_R$	m	m_R	m	1
422	$4m_R m_R$	$2m_R m_R$	$2m_R m_R$	m_R	m_R	m_R	1

	Zone-axis symmetries		
Point group	[001]	[*uv*0]	[*uvw*]
$4/m$	41_R	$2_R mm_R$	2_R
$\bar{4}$	4_R	m_R	1
4	4	m_R	1

	Zone-axis symmetries				
Point group	[001]	⟨100⟩	[*u*0*w*]	[*uv*0]	[*uvw*]
mmm	$2mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	2_R
$mm2$	$2mm$	$m1_R$	m	m_R	1
222	$2m_R m_R$	$2m_R m_R$	m_R	m_R	1

Table 2.5.3.4 (*cont.*)

Point group	Zone-axis symmetries		
	[010]	[*u*0*w*]	[*uvw*]
2/*m*	21_R	$2_R mm_R$	2_R
m	1_R	*m*	1
2	2	m_R	1

Point group	Zone-axis symmetry
	[*uvw*]
$\bar{1}$	2_R
1	1

the CBED symmetries for the two crystal (or incident-beam) settings which excite respectively the $+G$ and $-G$ reflections are drawn because the vertical rotation axes create the SMB patterns at different incident-beam orientations. [This had already been experienced for the case of symmetry 2_R (Goodman, 1975; Buxton *et al.*, 1976).] In the rectangular four-beam case, the symmetries for four settings which excite the $+G$, $+H$, $-G$ and $-H$ reflections are shown. For the diffraction groups $3m$, $3m_R$, $3m1_R$ and $6_R mm_R$, two different patterns are shown for the two crystal settings, which differ by $\pi/6$ rad from each other about the zone axis. Similarly, for the diffraction group $4_R mm_R$, two different patterns are shown for the two crystal settings, which differ by $\pi/4$ rad. Illustrations of these different symmetries are given in Fig. 2.5.3.7. The combination of the vertical threefold axis and a horizontal mirror plane introduces a new CBED symmetry 3_R. Similarly, the combination of the vertical sixfold rotation axis and an inversion centre introduces a new CBED symmetry 6_R.

There is an empirical and conventional technique for reproducing the symmetries of the SMB patterns which uses three operations of two-dimensional rotations, a vertical mirror at the centre of disc *O* and a rotation of π about the centre of a disc (1_R) without involving the reciprocal process. For example, we may consider 3_R between discs *F* and *F*′ in Table 2.5.3.5 in the case of diffraction group 31_R. Disc *F*′ is rotated anticlockwise not about the zone axis but *about the centre of disc O* by $2\pi/3$ rad (symbol 3) to coincide with disc *F*, and followed by a rotation of π rad (symbol *R*) about the centre of disc *F*′, resulting in the correct symmetry seen in Fig. 2.5.3.7. When the symmetries appearing between different SMB patterns are considered, this technique assumes that the symmetry operations are conducted after discs *O* and $\bar{O}$ are superposed. Another assumption is that the vertical mirror plane perpendicular to the line connecting discs *O* and $\bar{O}$ acts at the centre of disc *O* when the symmetries between two SMB patterns are considered. As an example, symmetry 3_R between discs *S* and $\bar{S}$ appearing in the two SMB patterns is reproduced by a threefold anticlockwise rotation of disc *S* about the centre of disc *O* (or $\bar{O}$) and followed by a rotation of π rad (*R*) about the centre of disc $\bar{S}$.

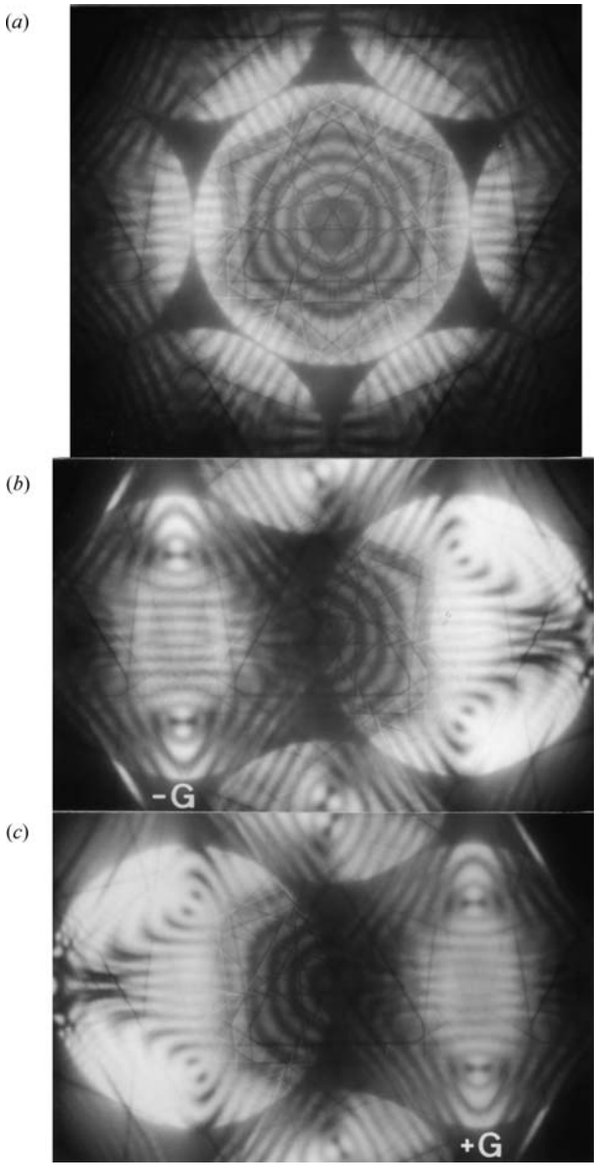


Fig. 2.5.3.5. CBED patterns of Si taken with the [111] incidence. (*a*) BP and WP show symmetry $3m_v$. (*b*) and (*c*) DPs show symmetry m_2 and DP symmetry $2_R m_{v'}$.

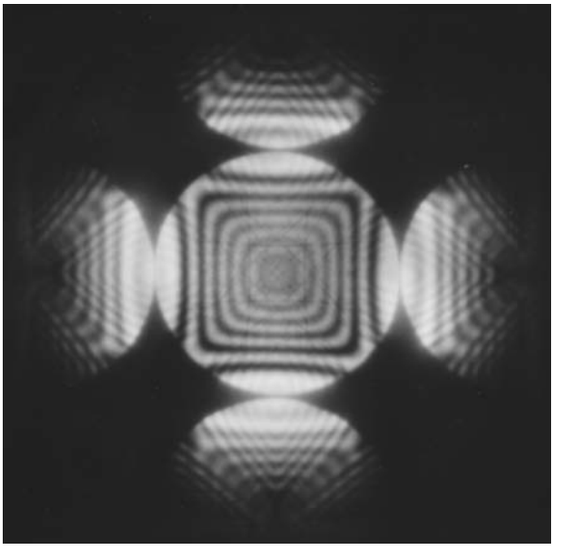


Fig. 2.5.3.6. CBED pattern of Si taken with the [100] incidence. The BP and WP show symmetry 4*mm*.

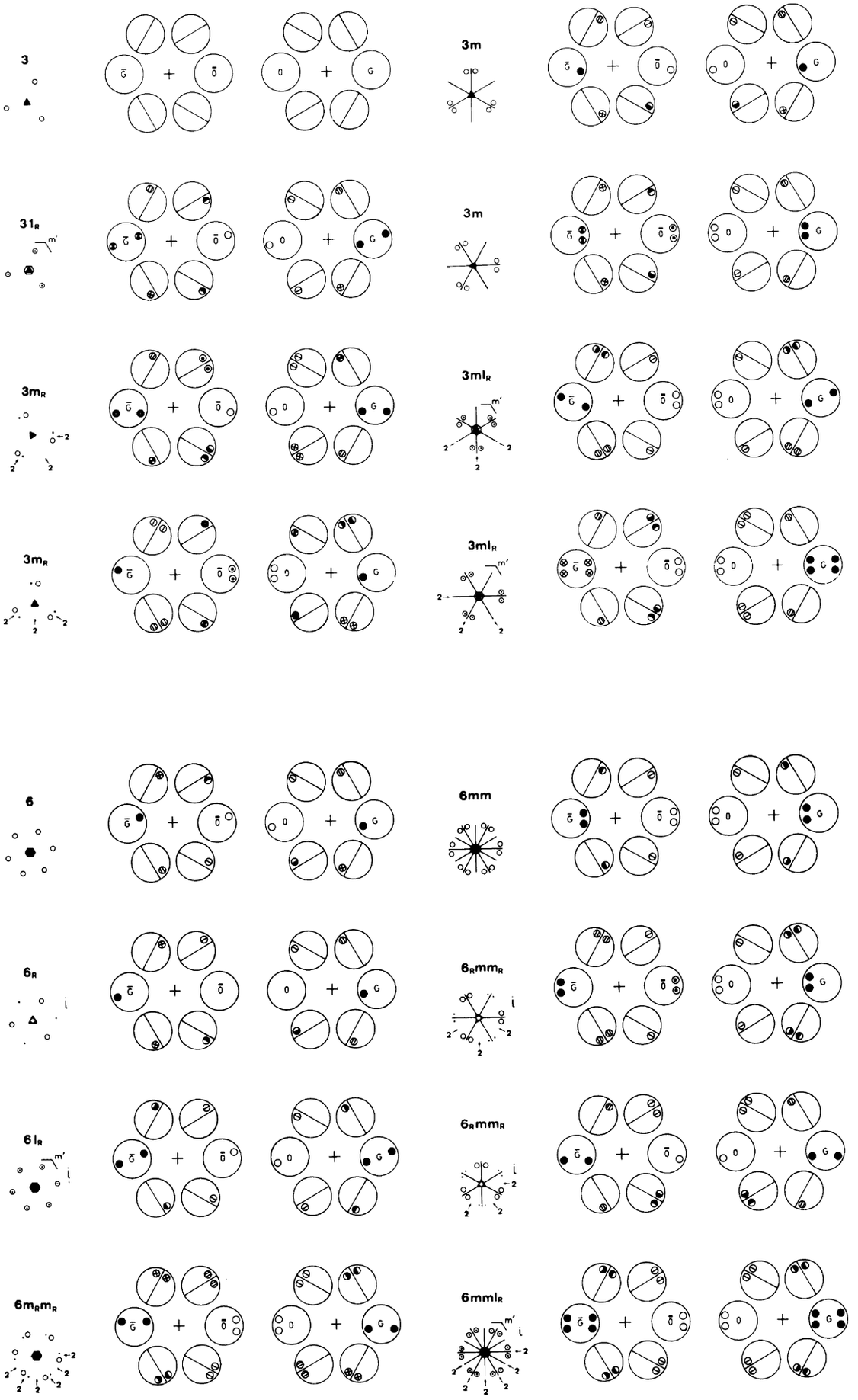


Fig. 2.5.3.7. Illustration of symmetries appearing in hexagonal six-beam, square four-beam and rectangular four-beam dark-field patterns expected for all the diffraction groups except for 1, 1_R, 2, 2_R and 21_R.

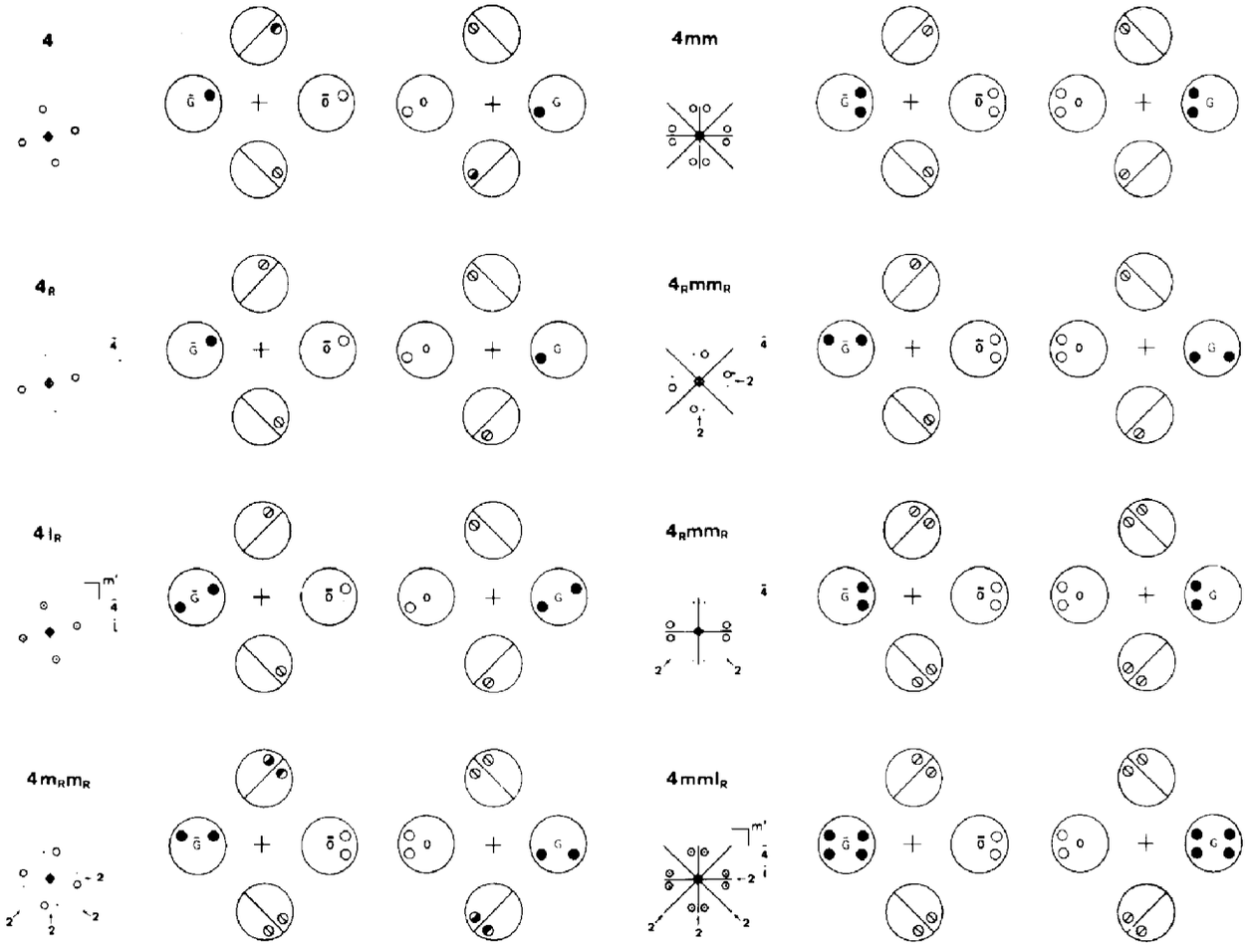


Fig. 2.5.3.7 (*cont.*).

Tables 2.5.3.5, 2.5.3.6 and 2.5.3.7 express the symmetries illustrated in Fig. 2.5.3.7 with the symmetry symbols for the hexagonal six-beam case, square four-beam case and rectangular four-beam case, respectively. In the fourth rows of the tables the symmetries of zone-axis patterns (BP and WP) are listed because combined use of the zone-axis pattern and the SMB pattern is efficient for symmetry determination. In the fifth row, the symmetries of the SMB pattern are listed. In the following rows, the symmetries appearing between the two SMB patterns are listed because the SMB symmetries appear not only in an SMB pattern but also in the pairs of SMB patterns. That is, for each diffraction group, all the possible SMB symmetries appearing in a pair of symmetric six-beam patterns, two pairs AB and AC of the square four-beam patterns and three pairs AB, AC and AD of the rectangular four-beam patterns are listed, though such pairs are not always needed for the determination of the diffraction groups. It is noted that the symmetries in parentheses are the symmetries which add no new symmetries, even if they are present. In the last row, the point groups which cause the diffraction groups listed in the first row are given.

By referring to Tables 2.5.3.5, 2.5.3.6 and 2.5.3.7, the characteristic features of the SMB method are seen to be as follows. CBED symmetry m_2 due to a horizontal twofold rotation axis can appear in every disc of an SMB pattern. Symmetry 1_R due to a horizontal mirror plane, however, appears only in disc G or H of an SMB pattern. In the hexagonal six-beam case, an inversion centre i produces CBED symmetry 6_R between discs S and S' due to the combination of an inversion centre and a vertical threefold rotation axis (and/or of a horizontal mirror plane and a vertical sixfold rotation axis). This indicates that one hexagonal six-beam pattern can reveal whether a specimen has an inversion centre or not, while the method of Buxton *et al.* (1976) requires two photographs for the inversion test. All the diffraction groups in Table 2.5.3.5 can be identified from one six-beam pattern except groups 3 and 6. Diffraction groups 3 and 6 cannot be distinguished from the hexagonal six-beam pattern because it is insensitive to the vertical axis. In the square four-beam case, fourfold rotary inversion $\bar{4}$ produces CBED symmetry 4_R between discs F and F' in one SMB pattern, while Buxton *et al.*'s method requires four photographs to identify fourfold rotary inversion. Although an inversion centre itself does not exhibit any symmetry in the square four-beam pattern, it causes symmetry 1_R due to the horizontal mirror plane produced by the combination of an inversion centre and the twofold rotation axis. Thus, symmetry 1_R is an indication of the existence of an inversion centre in the square four-beam case. All of the seven diffraction groups in Table 2.5.3.6 can be identified from one square four-beam pattern. One rectangular four-beam pattern can distinguish all the diffraction groups in Table 2.5.3.7 except the groups m and $2mm$. It is emphasized again that the inversion test can be carried out using one six-beam pattern or one square four-beam pattern.

Fig. 2.5.3.8 shows CBED patterns taken from a [111] pyrite (FeS_2) plate with an accelerating voltage of 100 kV. The space group of FeS_2 is $P2_1/a\bar{3}$. The diffraction group of the plate is 6_R due to a threefold rotation axis and an inversion centre. The zone-axis pattern of Fig. 2.5.3.8(a) shows threefold rotation symmetry in the BP and WP. The hexagonal six-beam pattern of Fig. 2.5.3.8(b) shows no symmetry higher than 1 in discs O, G, F and S but shows symmetry 6_R between discs S and S', which

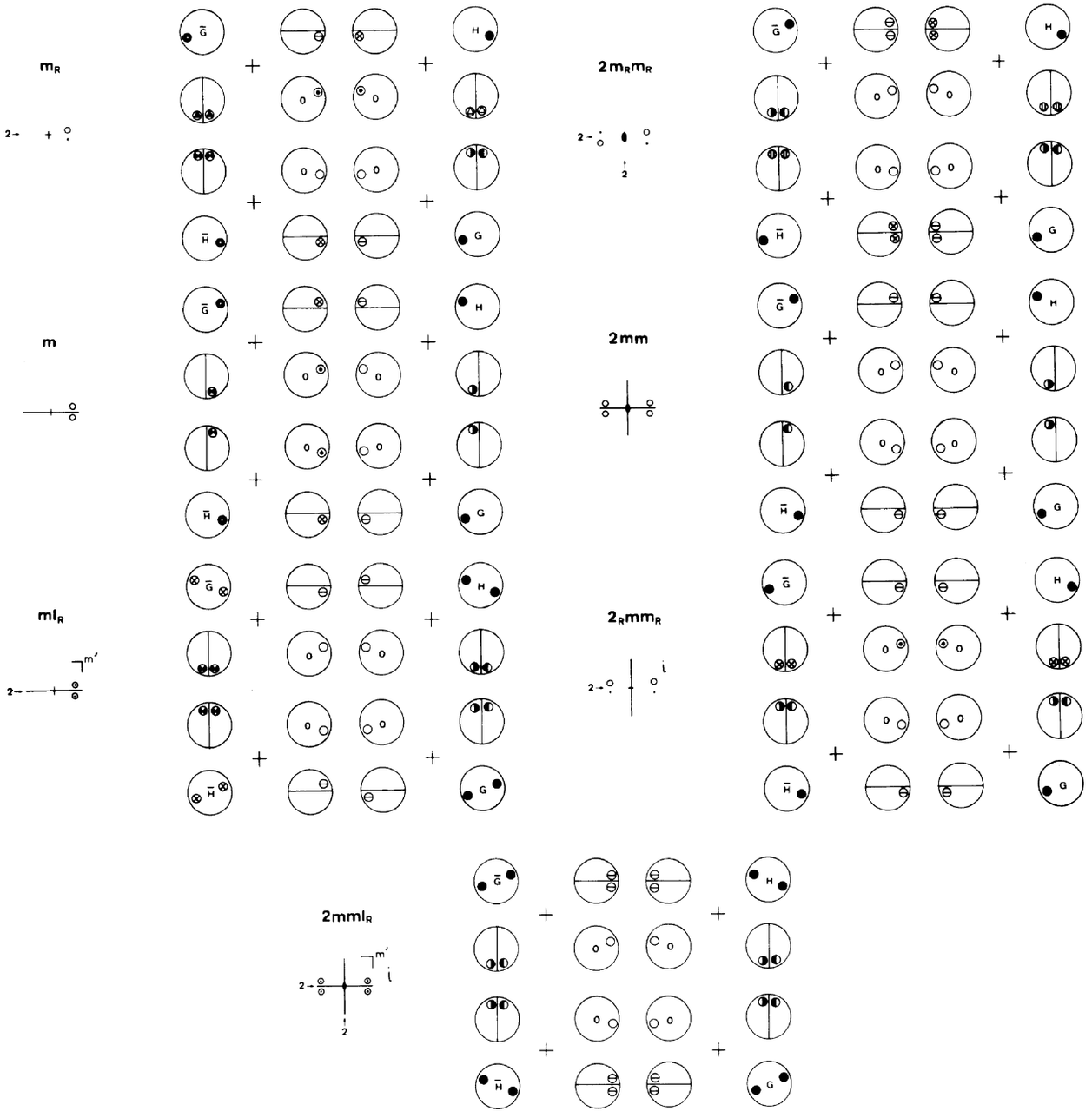


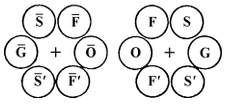
Fig. 2.5.3.7 (*cont.*).

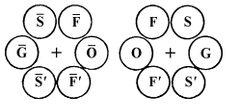
proves the existence of a threefold rotation axis and an inversion centre. The same symmetries are also seen in Fig. 2.5.3.8(*c*), where reflections $\bar{O}$, $\bar{G}$, $\bar{F}$, $\bar{S}$, $\bar{F}'$ and $\bar{S}'$ are excited. Table 2.5.3.5 indicates that diffraction group 6_R can be identified from only one hexagonal six-beam pattern, because no other diffraction groups give rise to the same symmetries in the six discs. When Buxton *et al.*'s method is used, three photographs or four patterns are necessary to identify diffraction group 6_R (see Table 2.5.3.3). In addition, if the symmetries between Figs. 2.5.3.8(*b*) and (*c*) are examined, symmetry 2_R between discs G and $\bar{G}$ and symmetry 6_R between discs F and $\bar{F}$ are found. All the experimental results agree exactly with the theoretical results given in Fig. 2.5.3.7 and Table 2.5.3.5.

Fig. 2.5.3.9 shows CBED patterns taken from a [110] V_3Si plate with an accelerating voltage of 80 kV. The space group of V_3Si is $Pm3n$. The diffraction group of the plate is $2mm1_R$ due to two vertical mirror planes and a horizontal mirror plane, a twofold rotation axis being produced at the intersection line of two perpendicular mirror planes. The zone-axis pattern of Fig. 2.5.3.9(*a*) shows symmetry $2mm$ in the BP and WP. The rectangular four-beam pattern of Fig. 2.5.3.9(*b*) shows symmetry 1_R in disc H due to the horizontal mirror plane and symmetry m_2 in both discs $\bar{S}$ and F' due to the twofold rotation axes in the [001] and [110] directions, respectively. The same symmetries are also seen in Fig. 2.5.3.9(*c*), where reflections $\bar{H}$, S' and $\bar{F}$ are excited. Table 2.5.3.7 implies that the diffraction group $2mm1_R$ can be identified from only one rectangular four-beam pattern, because no other diffraction groups give rise to the same symmetries in the four discs. When Buxton *et al.*'s method is used, two photographs or three patterns are necessary to identify diffraction group $2mm1_R$ (see Table 2.5.3.3). One can confirm the theoretically predicted symmetries between Fig. 2.5.3.9(*b*) and Fig. 2.5.3.9(*c*). All the experimental results agree exactly with the theoretical results given in Fig. 2.5.3.7 and Table 2.5.3.7.

These experiments show that the SMB method is quite effective for determining the diffraction group of slabs. Buxton *et al.*'s method identifies two-dimensional symmetry elements in the first place using a zone-axis pattern, and three-dimensional symmetry elements using DPs. On the other hand, the SMB method primarily finds many three-dimensional symmetry elements in an

Table 2.5.3.5. *Symmetries of hexagonal six-beam CBED patterns for diffraction groups*

		Projection diffraction group										
		31_R		$3m1_R$						61_R		
Diffraction group		3	31_R	$3m_R$		$3m$		$3m1_R$		6	6_R	61_R
Two-dimensional symmetry		3	3	3		$3m$		$3m$		6	3	6
Three-dimensional symmetry			m'	$2'$				m', $(2')$			i	m', (i)
Zone-axis pattern	Bright-field pattern	3	6	$3m$		$3m$		$6mm$		6	3	6
	Whole-field pattern	3	3	3		$3m$		$3m$		6	3	6
Hexagonal six-beam pattern	O	1	1	1	m_2	1	m_v	m_2	m_v	1	1	1
	G	1	1_R	m_2	1	1	m_v	1_R	$1_R m_v(m_2)$	1	1	1_R
	F	1	1	m_2	1	1	1	1	m_2	1	1	1
	S	1	1	1	m_2	1	1	m_2	1	1	1	1
	FF'	1	3_R	1	1	1	m_v	3_R	$3_R m_v$	1	1	3_R
	SS'	1	1	1	1	1	m_v	1	m_v	1	6_R	6_R
A pair of symmetrical six-beam patterns 	$\pm O$	1	1_R	m_2	1	m_v	1	$m_v 1_R$	$1_R m_2$	2	1	$2(1_R)$
	$\pm G$	1	1	1	m_R	m_v	1	$m_v m_R$	1	2	2_R	21_R
	$\pm F$	1	1	1	1	m_v	1	m_v	1	1	6_R	6_R
	$\pm S$	1	3_R	1	1	m_v	1	$3_R m_v$	3_R	1	1	3_R
	$F'\bar{F}$	1	1	1	m_R	1	1	m_R	1	2	1	2
	$S'\bar{S}$	1	1	m_R	1	1	1	1	m_R	2	1	2
Point group		23, 3	$\bar{6}$	432, 32		$\bar{4}3m$, $3m$		$\bar{6}m2$		6	$m3$, 3	$6/m$

		Projection diffraction group				
		$6mm1_R$				
Diffraction group		$6m_R m_R$	$6mm$	$6_R mm_R$		$6mm1_R$
Two-dimensional symmetry		6	$6mm$	$3m$		$6mm$
Three-dimensional symmetry		$2'$		i, $(2')$		m', $(i, 2')$
Zone-axis pattern	Bright-field pattern	$6mm$	$6mm$	$3m$		$6mm$
	Whole-field pattern	6	$6mm$	$3m$		$6mm$
Hexagonal six-beam pattern	O	m_2	m_v	1	$m_v(m_2)$	$m_v(m_2)$
	G	m_2	m_v	m_2	m_v	$1_R m_v(m_2)$
	F	m_2	1	m_2	1	m_2
	S	m_2	1	1	m_2	m_2
	FF'	1	m_v	1	m_v	$3_R m_v$
	SS'	1	m_v	6_R	$6_R m_v$	$6_R m_v$
A pair of symmetrical six-beam patterns 	$\pm O$	$2m_2$	$2m_{v'}$	$m_v(m_2)$	1	$2(1_R)m_{v'}(m_2)$
	$\pm G$	$2m_R$	$2m_{v'}$	$2_R m_v$	$2_R m_R$	$21_R m_{v'}(m_R)$
	$\pm F$	1	$m_{v'}$	$6_R m_v$	6_R	$6_R m_{v'}$
	$\pm S$	1	$m_{v'}$	m_v	1	$3_R m_{v'}$
	$F'\bar{F}$	$2m_R$	2	1	m_R	$2m_R$
	$S'\bar{S}$	$2m_R$	2	m_R	1	$2m_R$
Point group		622	$6mm$	$m3m$, $\bar{3}m$		$6/mmm$

SMB pattern, and two-dimensional symmetry elements from a pair of SMB patterns, as shown in Tables 2.5.3.5, 2.5.3.6 and 2.5.3.7. Therefore, the use of a ZAP and SMB patterns is the most efficient way to find as many crystal symmetry elements in a specimen as possible.

2.5.3.3. Space-group determination

2.5.3.3.1. Lattice-type determination

When the point group of a specimen crystal is determined, the crystal axes may be found from a spot diffraction pattern recorded at a high-symmetry zone axis, using the orientations of the symmetry elements determined in the course of point-group determination. Integral-number indices are assigned to the spots of the diffraction patterns. The systematic absence of reflections indicates the lattice type of the crystal. It should be noted that reflections forbidden by the lattice type are always absent, even if dynamical diffraction takes place. (This is true for all sample thicknesses and accelerating voltages.) By comparing the experimentally obtained absences and the extinction rules known for the lattice types [P, C (A, B), I, F and R], a lattice type may be identified for the crystal concerned.

2.5.3.3.2. Identification of screw axes and glide planes

There are three space-group symmetry elements of diperiodic plane figures: (1) a horizontal screw axis $2'_1$, (2) a vertical glide plane g with a horizontal glide vector and (3) a horizontal glide plane g'. These are related to the point-group symmetry elements $2'$, m and m' of diperiodic plane figures, respectively. (It is noted that these symmetry elements and ten point-group symmetry elements form 80 space groups.)

Table 2.5.3.6. *Symmetries of square four-beam CBED patterns for diffraction groups*

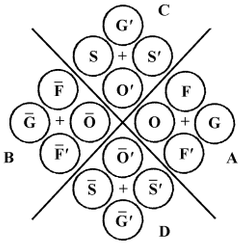
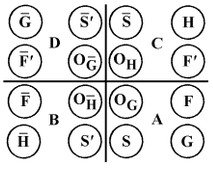
			Projection diffraction group							
			41_R			$4mm1_R$				
Diffraction group			4	4_R	41_R	$4m_Rm_R$	$4mm$	4_Rmm_R		$4mm1_R$
Two-dimensional symmetry			4	(2)	4	4	$4mm$	($2mm$)		$4mm$
Three-dimensional symmetry				$\bar{4}$	m', (i, $\bar{4}$)	$2'$		$\bar{4}$, $2'$		m', (i, $2'$, $\bar{4}$)
Zone-axis pattern	Bright-field pattern		4	4	4	$4mm$	$4mm$	$4mm$		$4mm$
	Whole-field pattern		4	2	4	4	$4mm$	$2mm$		$4mm$
Square four-beam pattern		O	1	1	1	m_2	m_v	m_2	m_v	$m_v(m_2)$
		G	1	1	1_R	m_2	m_v	m_2	m_v	$1_Rm_v(m_2)$
		F	1	1	1	m_2	1	1	m_2	m_2
		FF'	1	4_R	4_R	1	m_v	4_R	4_Rm_v	4_Rm_v
Two pairs of square four-beam patterns 	AB	$\pm O$	2	2	$2(1_R)$	$2m_2$	$2m_{v'}$	$2m_2$	$2m_{v'}$	$2(1_R)m_{v'}(m_2)$
		$\pm G$	2	2	21_R	$2m_R$	$2m_{v'}$	$2m_R$	$2m_{v'}$	$21_Rm_{v'}(m_R)$
		FF'	2	2	2	$2m_R$	2	2	$2m_R$	$2m_R$
		$\pm F$	1	4_R	4_R	1	$m_{v'}$	4_R	$4_Rm_{v'}$	$4_Rm_{v'}$
	AC	OO'	4	4	4	$4m_2$	$4m_{v''}$	$4m_v$	$4m_2$	$4m_{v''}(m_2)$
		GG'	4	4_R	41_R	$4m_R$	$4m_{v''}$	4_Rm_v	4_Rm_R	$41_Rm_{v''}(m_R)$
		FS	4	1	4	$4m_R$	4	m_R	1	$41_Rm_{v''}(m_R)$
		FS'	1	1	1_R	1	$m_{v''}$	m_v	1	$1_Rm_{v''}$
Point group			4	$\bar{4}$	$4/m$	432, 422	$4mm$	$\bar{4}3m$, $\bar{4}2m$		$m3m$, $4/mmm$

Table 2.5.3.7. *Symmetries of rectangular four-beam CBED patterns for diffraction groups*

			Projection diffraction group						
			$m1_R$			$2mm1_R$			
Diffraction group			m_R	m	$m1_R$	$2m_Rm_R$	$2mm$	2_Rmm_R	$2mm1_R$
Two-dimensional symmetry				m	m	2	$2mm$	m	$2mm$
Three-dimensional symmetry			$2'$		m', $2'$	$2'$		$2'$, i	m', $2'$, i
Zone-axis pattern	Bright-field pattern		m	m	$2mm$	$2mm$	$2mm$	m	$2mm$
	Whole-field pattern		1	m	m	2	$2mm$	m	$2mm$
Rectangular four-beam pattern		O	1	1	1	1	1	1	1
		G	1	1	1_R	1	1	1	1_R
		F	m_2	1	m_2	m_2	1	m_2	m_2
		S	1	1	1	m_2	1	1	m_2
Three pairs of rectangular four-beam patterns 	AB	$O_GO_{\bar{H}}$	m_2	1	m_2	m_2	m_v	$m_v(m_2)$	$m_v(m_2)$
		$G\bar{H}$	1	1	1	m_R	m_v	m_v	m_vm_R
		$F\bar{F}$	1	1	1	1	m_v	2_Rm_v	2_Rm_v
		SS'	1	1	1_R	1	m_v	m_v	m_v1_R
	AC	O_GO_H	1	m_v	m_v	m_2	$m_{v'}$	1	$m_{v'}(m_2)$
		GH	m_R	m_v	m_vm_R	m_R	$m_{v'}$	m_R	$m_{v'}m_R$
		FF'	1	m_v	m_v1_R	1	$m_{v'}$	1	$m_{v'}1_R$
		$S\bar{S}$	1	m_v	m_v	1	$m_{v'}$	2_R	$2_Rm_{v'}$
	AD	$O_GO_{\bar{G}}$	1	1	1_R	2	2	1	$2(1_R)$
		GG	1	1	1	2	2	2_R	21_R
		$F\bar{F}'$	1	1	1	$2m_R$	2	1	$2m_R$
		$S\bar{S}'$	m_R	1	m_R	$2m_R$	2	m_R	$2m_R$
Point group			2, 222, $mm2$, 4, $\bar{4}$, 422, $4mm$, $\bar{4}2m$, 32, $\bar{6}$, 622, $6mm$, $\bar{6}m2$, 23, 432, $\bar{4}3m$	m, $mm2$, $4mm$, $\bar{4}2m$, $3m$, $\bar{6}$, $6mm$, $\bar{6}m2$, $\bar{4}3m$	$mm2$, $4mm$, $42m$, $6mm$, $\bar{6}m2$, $\bar{4}3m$	222, 422, $\bar{4}2m$, 622, 23, 432	$mm2$, $\bar{6}m2$	$2/m$, mmm, $4/m$, $4/mmm$, $\bar{3}m$, $\bar{6}/m$, $6/mmm$, $m3$, $m3m$	mmm, $4/mmm$, $m3$, $m3m$, $6/mmm$

The ordinary extinction rules for screw axes and glide planes hold only in the approximation of kinematical diffraction. The kinematically forbidden reflections caused by these symmetry elements appear owing to *Umweganregung* of dynamical diffraction. Extinction of intensity, however, does take place in these reflections at certain crystal orientations with respect to the incident beam (*i.e.* in certain regions within a CBED disc). This dynamical extinction was first predicted by Cowley & Moodie (1959) and was discussed by Miyake *et al.* (1960) and Cowley *et al.* (1961). Goodman & Lehmpfuhl (1964) first observed the dynamical extinction as dark cross lines in kinematically forbidden reflection discs of CBED patterns of CdS. Gjønnes & Moodie

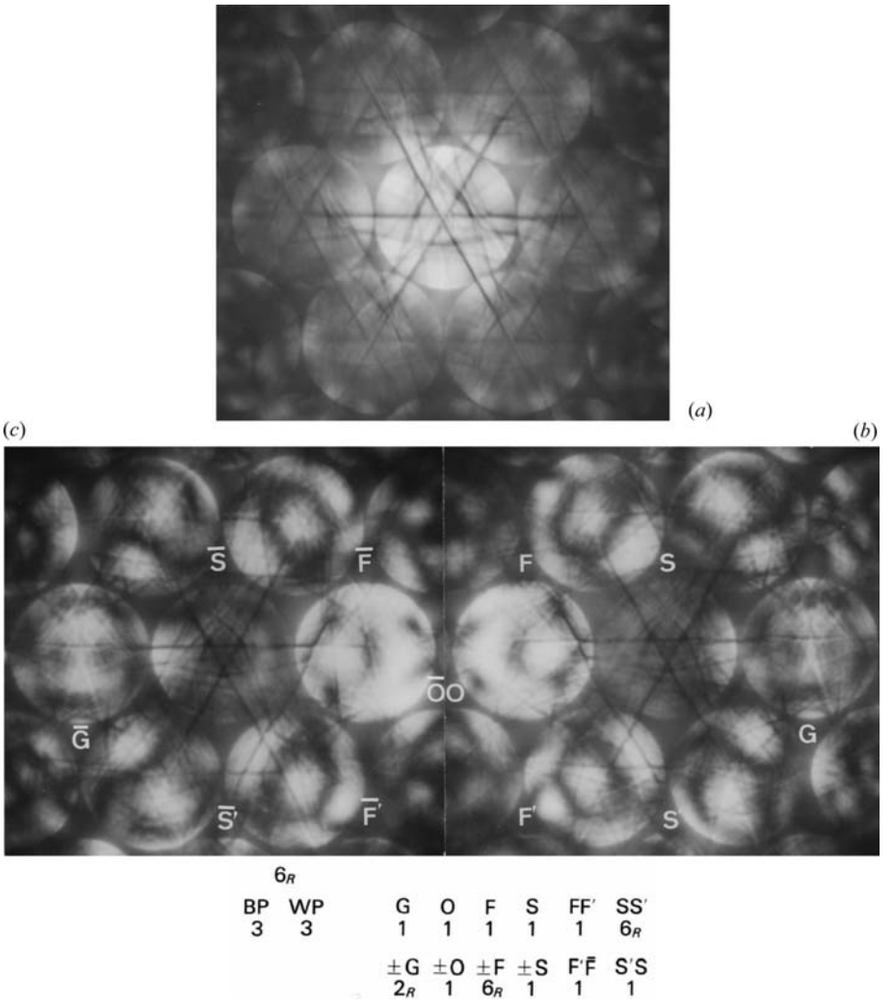


Fig. 2.5.3.8. CBED patterns of FeS_2 taken with the [111] incidence. (*a*) Zone-axis pattern, (*b*) hexagonal six-beam pattern with excitation of reflection $+G$, (*c*) hexagonal six-beam pattern with excitation of reflection $-G$. Symmetry 6_R is noted between discs S and S' and discs $\bar{F}$ and $\bar{F}'$.

(1965) discussed the dynamical extinction in a more general way considering not only ZOLZ reflections but also HOLZ reflections. They completely clarified the dynamical extinction rules by considering the exact cancellation which may occur along certain symmetry-related multiple-scattering paths. Based on the results of Gjønnes & Moodie (1965), Tanaka, Sekii & Nagasawa (1983) tabulated the dynamical extinctions expected at all the possible crystal orientations for all the space groups. These were later tabulated in a better form on pages 162 to 172 of the book by Tanaka & Terauchi (1985).

Fig. 2.5.3.10(*a*) illustrates *Umweganregung* paths to a kinematically forbidden reflection. The $0k0$ (k = odd) reflections are kinematically forbidden because a b-glide plane exists perpendicular to the a axis and/or a 2_1 screw axis exists in the b direction. Let us consider an *Umweganregung* path a in the zeroth-order Laue zone to the 010 forbidden reflection and path b which is symmetric to path a with respect to axis k. Owing to the translation of one half of the lattice parameter b caused by the glide plane and/or the 2_1 screw axis, the following relations hold between the crystal structure factors:

$$F(h,k) = F(\bar{h},k) \quad \text{for } k = 2n,$$
$$F(h,k) = -F(\bar{h},k) \quad \text{for } k = 2n+1. \qquad (2.5.3.1)$$

That is, the structure factors of reflections $hk0$ and $\bar{h}k0$ have the same phase for even k but have opposite phases for odd k.

Since an *Umweganregung* path to the kinematically forbidden reflection $0k0$ contains an odd number of reflections with odd k, the following relations hold:

$$F(h_1,k_1)F(h_2,k_2)\ldots F(h_n,k_n) \quad \text{for path } a$$
$$= -F(\bar{h}_1,k_1)F(\bar{h}_2,k_2)\ldots F(\bar{h}_n,k_n) \quad \text{for path } b, \qquad (2.5.3.2)$$

where

$$\sum_{i=1}^{n} h_i = 0, \quad \sum_{i=1}^{n} k_i = k \ (k = \text{odd})$$

and the functions including the excitation errors are omitted because only the cases in which the functions are the same for all the paths are considered. The excitation errors for paths a and b become the same when the projection of the Laue point along the zone axis concerned, L, lies on axis k. Since the two waves passing through paths a and b have the same amplitude but opposite signs, these waves are superposed on the $0k0$ discs (k = odd) and cancel out, resulting in dark lines A in the forbidden discs, as shown in Fig. 2.5.3.10(*b*). The line A runs parallel to axis k passing through the projection point of the zone axis.

In path c, the reflections are arranged in the reverse order to those in path b. When the 010 reflection is exactly excited, two paths a and c are symmetric with respect to the bisector m'–m' of

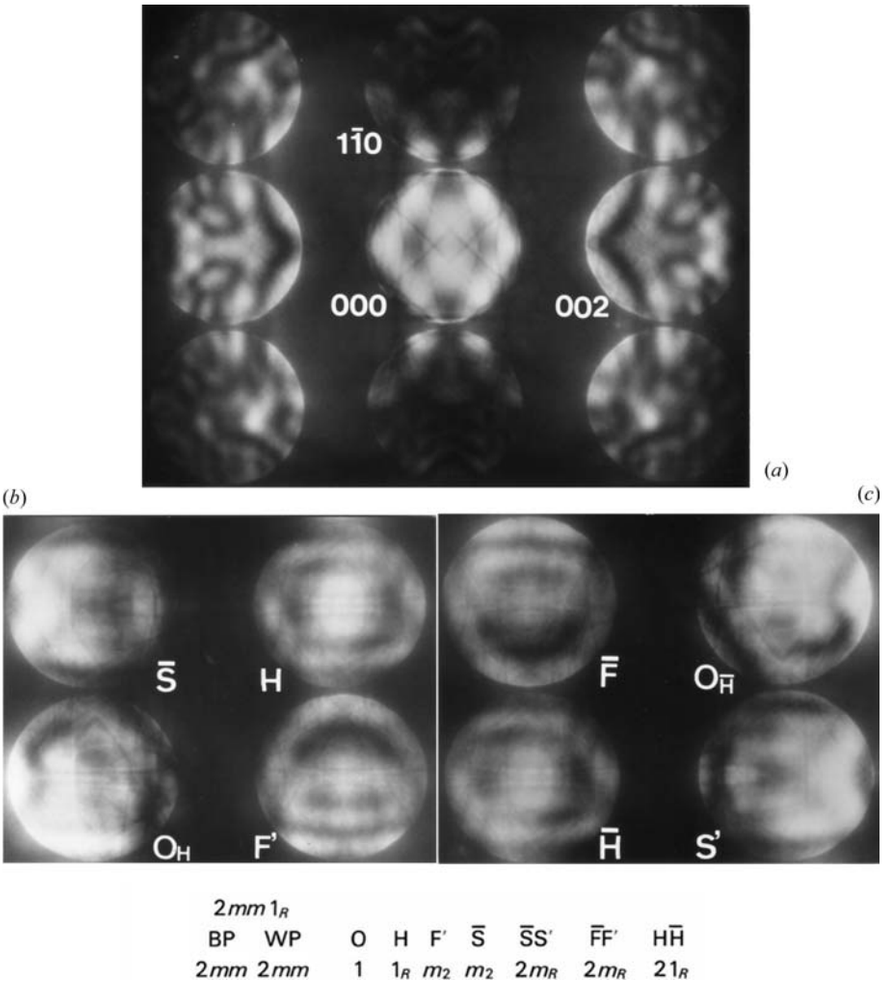


Fig. 2.5.3.9. CBED patterns of V_3Si taken with the [110] incidence. (*a*) Zone-axis pattern, (*b*) rectangular four-beam pattern with excitation of reflections H, $\bar{S}$ and F, (*c*) rectangular four-beam pattern with excitation of reflections $\bar{H}$, S and $\bar{F}$.

the 010 vector having the same excitation errors. The following equation holds:

$$F(h_1, k_1)F(h_2, k_2)\ldots F(h_n, k_n) \quad \text{for path } a$$
$$= -F(\bar{h}_n, k_n)F(\bar{h}_{n-1}, k_{n-1})\ldots F(\bar{h}_1, k_1) \quad \text{for path } c. \tag{2.5.3.3}$$

Since the waves passing through these paths have the same amplitude but opposite signs, these waves are superposed on the 010 discs and cancel out, resulting in dark line B in this disc, as shown in Fig. 2.5.3.10(*b*). Line B appears perpendicular to line A at the exact Bragg positions. When *Umweganregung* paths are present only in the zeroth-order Laue zone, the glide plane and screw axis produce the same dynamical extinction lines A and B. We call these lines A_2 and B_2 lines, subscript 2 indicating that the *Umweganregung* paths lie in the zeroth-order Laue zone.

The dynamical extinction effect is analogous to interference phenomena in the Michelson interferometer. That is, the incident beam is split into two beams by Bragg reflections in a crystal. These beams take different paths, in which they suffer a relative phase shift of π and are finally superposed on a kinematically forbidden reflection to cancel out.

When the paths include higher-order Laue zones, the glide plane produces only extinction lines A but the screw axis causes only extinction lines B. These facts are attributed to the different relations between structure factors for a 2_1 screw axis and a glide plane,

$$F(hkl) = (-1)^k F(\bar{h}k\bar{l}) \quad \text{for a } 2_1 \text{ screw axis in the [010] direction,} \tag{2.5.3.4}$$

$$F(hkl) = (-1)^k F(\bar{h}kl) \quad \text{for a } b \text{ glide in the (100) plane.} \tag{2.5.3.5}$$

In the case of the glide plane, extinction lines A are still formed because two waves passing through paths a and b have opposite signs to each other according to equation (2.5.3.5), but extinction lines B are not produced because equation (2.5.3.4) holds only for the 2_1 screw axis. In the case of the 2_1 screw axis, only the waves passing through paths a and c have opposite signs according to equation (2.5.3.4), forming extinction lines B only. We call these lines A_3 and B_3 dynamical extinction lines, suffix 3 indicating the *Umweganregung* paths being *via* higher-order Laue zones.

It was predicted by Gjønnes & Moodie (1965) that a horizontal glide plane g' gives a dark spot at the crossing point between extinction lines A and B (Fig. 2.5.3.10*b*) due to the cancellation between the waves passing through paths b and c. Tanaka, Terauchi & Sekii (1987) observed this dynamical extinction, though it appeared in a slightly different manner to that predicted by Gjønnes & Moodie (1965). Table 2.5.3.8 summarizes the appearance of the dynamical extinction lines for the glide planes g and g' and the 2_1 screw axis. The three space-group symmetry elements can be identified from the observed extinctions because these three symmetry elements produce different kinds of dynamical extinctions.

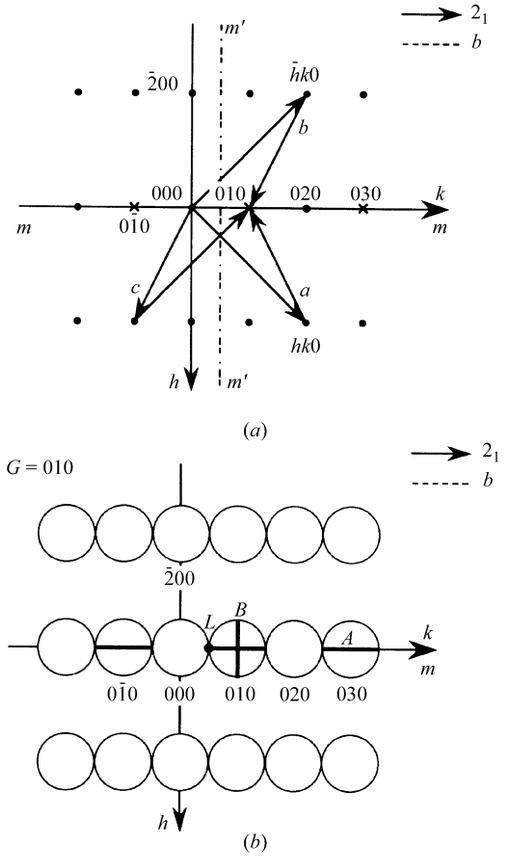


Fig. 2.5.3.10. Illustration of the production of dynamical extinction lines in kinematically forbidden reflections due to a b-glide plane and a 2_1 screw axis. (a) *Umweganregung* paths a, b and c. (b) Dynamical extinction lines A are formed in forbidden reflections $0k0$ (k = odd). Extinction line B perpendicular to the lines A is formed in the exactly excited 010 reflection.

In principle, a horizontal screw axis and a vertical glide plane can be distinguished by observations of the extinction lines A_3 and B_3. It is, however, not easy to observe the extinction lines A_3 and B_3 because broad extinction lines A_2 and B_2 appear at the same time. The presence of the extinction lines A_3 and B_3 can be revealed by inspecting the symmetries of fine defect HOLZ lines appearing in the forbidden reflections instead of by direct observation of the lines A_3 and B_3 (Tanaka, Sekii & Nagasawa, 1983). That is, if HOLZ lines form lines A_3 and B_3, HOLZ lines are symmetric with respect to the extinction lines A_2 and B_2. If HOLZ lines do not form lines A_3 and B_3, HOLZ lines are asymmetric with respect to the extinction lines A_2 and B_2. When the HOLZ lines are symmetric about the extinction lines A_2, the specimen crystal has a glide plane. When the HOLZ lines are symmetric with respect to lines B_2, a 2_1 screw axis exists. It should

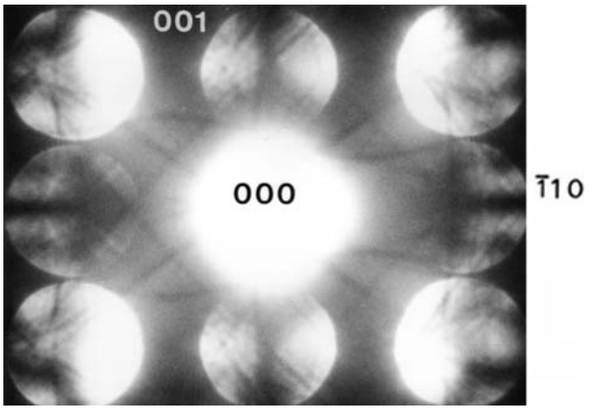


Fig. 2.5.3.12. CBED pattern of FeS_2 taken with the [110] electron-beam incidence. In the 001 and $00\bar{1}$ discs, HOLZ lines are asymmetric with respect to extinction lines A_2, indicating the existence of a 2_1 screw axis parallel to the c axis. In the $\bar{1}10$ and $1\bar{1}0$ discs, HOLZ lines are symmetric with respect to extinction lines A_2, indicating existence of a glide plane perpendicular to the c axis.

be noted that a relatively thick specimen area has to be selected to observe HOLZ lines in ZOLZ reflection discs.

Fig. 2.5.3.11 shows CBED patterns taken from (a) thin and (b) thick areas of FeS_2, whose space group is $P2_1/a\bar{3}$, at the 001 Bragg setting with the [100] electron-beam incidence. In the case of the thin specimen (Fig. 2.5.3.11a), only the broad dynamical extinction lines formed by the interaction of ZOLZ reflections are seen in the odd-order discs. On the other hand, fine HOLZ lines are clearly seen in the thick specimen (Fig. 2.5.3.11b). The HOLZ lines are symmetric with respect to both A_2 and B_2 extinction lines. This fact proves the presence of the extinction lines A_3 and B_3, or both the c glide in the (010) plane and the 2_1 screw axis in the c direction, this fact being confirmed by consulting Table 2.5.3.9. Fig. 2.5.3.12 shows a [110] zone-axis CBED pattern of FeS_2. A_2 extinction lines are seen in both the 001 and $\bar{1}10$ discs. Fine HOLZ lines are symmetric with respect to the A_2 extinction lines in the $\bar{1}10$ disc but asymmetric about the A_2 extinction line in the 001 disc, indicating formation of the A_3 extinction line only in the $\bar{1}10$ disc. This proves the existence of a 2_1 screw axis in the [001] direction and an a glide in the (001) plane. The appearance of HOLZ lines is easily changed by a change of a few hundred volts in the accelerating voltage. Steeds & Evans (1980) demonstrated for spinel changes in the appearance of HOLZ lines in the ZOLZ discs at accelerating voltages around 100 kV.

Another practical method for distinguishing between glide planes and 2_1 screw axes is that reported by Steeds *et al.* (1978). The method is based on the fact that the extinction lines are observable even when a crystal is rotated with a glide plane kept parallel and with a 2_1 screw axis perpendicular to the incident

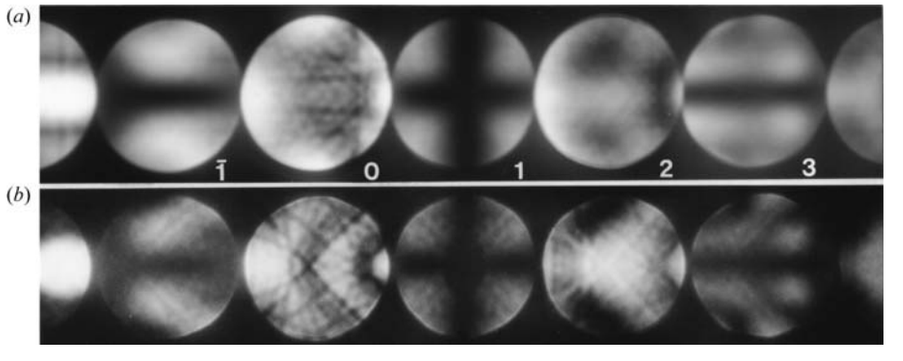


Fig. 2.5.3.11. CBED patterns obtained from (a) thin and (b) thick areas of (001) FeS_2. (a) Dynamical extinction lines A_2 and B_2 are seen. (b) Extinction lines A_3 and B_3 as well as A_2 and B_2 are formed because HOLZ lines are symmetric about lines A_2 and B_2.

Table 2.5.3.8. *Dynamical extinction rules for an infinitely extended parallel-sided specimen*

Symmetry elements of plane-parallel specimen	Orientation to specimen surface	Dynamical extinction lines	
		Two-dimensional (zeroth Laue zone) interaction	Three-dimensional (HOLZ) interaction
Glide planes	perpendicular: g	A_2 and B_2	A_3
	parallel: g'	—	intersection of A_3 and B_3
Twofold screw axis	perpendicular: 2_1	—	—
	parallel: $2_1'$	A_2 and B_2	B_3

beam. With reference to Fig. 2.5.3.10(*a*), extinction lines A_3 produced by a glide plane remain even when the crystal is rotated with respect to axis h but the lines are destroyed by a rotation of the crystal about axis k. Extinction lines B_3 originating from a 2_1 screw axis are not destroyed by a crystal rotation about axis k but the lines are destroyed by a rotation with respect to axis h.

2.5.3.3.3. Space-group determination

We now describe a space-group determination method which uses the dynamical extinction lines caused by the horizontal screw axis $2_1'$ and the vertical glide plane g of an infinitely extended parallel-sided specimen. We do not use the extinction due to the glide plane g' because observation of the extinction requires a laborious experiment. It should be noted that a vertical glide plane with a glide vector not parallel to the specimen surface cannot be a symmetry element of a specimen of finite thickness; however, the component of the glide vector perpendicular to the incident beam acts as a symmetry element g. (Which symmetry elements are observed by CBED is discussed in Section 2.5.3.3.5.) The 2_1, 4_1, 4_3, 6_1, 6_3 and 6_5 screw axes of crystal space groups that are set perpendicular to the incident beam act as a symmetry element $2_1'$ because two or three successive operations of 4_1, 4_3, 6_1, 6_3 and 6_5 screw axes make them equivalent to a 2_1 screw axis: $(4_1)^2 = (4_3)^2 = (6_1)^3 = (6_3)^3 = (6_5)^3 = 2_1$. The 4_2, 3_1, 3_2, 6_2 and 6_4 screw axes that are set perpendicular to the incident beam do not produce dynamical extinction lines because the 4_2 screw axis acts as a twofold rotation axis due to the relation $(4_2)^2 = 2$, the 3_1 and 3_2 screw axes give no specific symmetry in CBED patterns, and the 6_2 and 6_4 screw axes are equivalent to 3_1 and 3_2 screw axes due to the relations $(6_2)^2 = 3_2$ and $(6_4)^2 = 3_1$. Modifications of the dynamical extinction rules were investigated by Tanaka, Sekii & Nagasawa (1983) when more than one crystal symmetry element (that gives rise to dynamical extinction lines) coexists and when the symmetry elements are combined with various lattice types. Using these results, dynamical extinction lines A_2, A_3, B_2 and B_3 expected from all the possible crystal settings for all the space groups were tabulated.

Table 2.5.3.9 shows all the dynamical extinction lines appearing in the kinematically forbidden reflections for all the possible crystal settings of all the space groups. The first column gives space groups. In each of the following pairs of columns, the left-hand column of the pair gives the reflection indices and the symmetry elements causing the extinction lines and the right-hand column gives the types of the extinction lines. The (second) suffixes 1, 2 and 3 of a 2_1 screw axis in each column distinguish the first, the second and the third screw axis of the space group (as in the symbols 2_{11} and 2_{12} of space group $P2_12_12$). The glide symbols in the [001] column for space group $P4/nnc$ have two suffixes (n_{21} and n_{22}). The first suffix 2 denotes the second glide plane of the space group. The second suffixes 1 and 2, which appear in the tetragonal and cubic systems, distinguish two equivalent glide planes which lie in the x and y planes. The equivalent planes are distinguished only for the cases of [100], [010] and [001] electron-beam incidences, for convenience. The c-glide planes of space group $Pcc2$ are distinguished with symbols c_1 and c_2 (the first suffix only), because the equivalent planes do not exist. The glide symbol in the [001] column for space group $P4/mbm$ has only one suffix 1 or 2. The suffix distinguishes the equivalent glide planes lying in the x and y planes. The first suffix to distinguish the first and the second glide planes is not necessary because the space group has only one glide symbol b. When the index of the incident-beam direction is expressed with a symbol like $[h0l]$ for point groups 2, m and $2/m$, the index h or l can take a value of zero. That is, the extinction rules are applicable to the [100] and [001] electron-beam incidences. However, if columns for [100], [010] and [001] incidences are present, as in the case of point group $mm2$, $[hk0]$, $[0kl]$ and $[h0l]$ incidences are only for nonzero h, k and l. The reflections in which the extinction lines appear are always perpendicular to the corresponding incident-beam directions ($0k'l' \perp [0kl]$, $h'k'0 \perp [hk0]$, ...). The indices of the reflections in which extinction lines appear are odd if no remark is given. For c-glide planes of space groups $R3c$ and $R\bar{3}c$ and for d-glide planes, the reflections in which extinction lines appear are specified as $6n + 3$ and $4n + 2$ orders, respectively.

The number of indistinguishable space groups was first counted by Tanaka, Sekii & Nagasawa (1983) but later corrections were made by Eades & Spence (1987). It was found that 177 space groups out of 230 can be identified using the extinction lines (Tanaka *et al.*, 2002). Another reference for space-group determination is due to Eades (1988). The indistinguishable space-group sets using the extinction lines are listed in Table 2.5.3.10. Most of the sets are caused by the fact that CBED cannot identify 4_2, 3_1 (3_2) and 6_2 (6_4) screw axes. However, these sets can be rather easily distinguished in the ordinary way, that is, by observing how the intensities of the reflections which may be kinematically forbidden change when the crystal orientation is varied. If the axis concerned is a screw axis, kinematically forbidden reflections show a sudden decrease in intensity when an orientation change causes the loss of *Umweganregung* paths. If the axis is a rotation axis, the intensities of the reflections do not change conspicuously for such an orientation change. Using this test, each space group in the 23 sets can be identified except the pairs in parentheses and pairs (16) and (17) in Table 2.5.3.10 (see Eades, 1988).

Tsuda *et al.* (2000) showed theoretically that the coherent CBED method can distinguish between space groups ($I23$ and $I2_13$) and between ($I222$ and $I2_12_12_1$), which are indistinguishable pairs (16) and (17), respectively, in Table 2.5.3.10. The coherent CBED pattern is obtained in such a way that the convergence angle of the incident beam is set to a larger value than usual to make adjacent CBED discs overlap (Dowell & Goodman, 1973). When the focus point is displaced from the specimen, or a certain area is illuminated, sinusoidal interference fringes of the lattice spacing corresponding to the adjacent discs are formed in the overlapping regions if the probe size of the incident beam is smaller than the lattice spacing. (If the focus point of the incident beam is on the specimen, each overlapping region of the CBED discs shows uniform intensity.) Formation of the interference fringes was explained in detail first by Spence & Cowley (1978). Vine *et al.* (1992) showed distortion-free interference fringes from 6*H*-SiC and succeeded in observing the fringes with a shift of half a period due to a glide plane. Tsuda *et al.*'s method

Table 2.5.3.9. *Dynamical extinction lines appearing in ZOLZ reflections for all crystal space groups except Nos. 1 and 2*

Point groups 2, *m*, 2/*m* (second setting, unique axis *b*)

Space group		Incident-beam direction [*h*0*l*]
3	*P*2	
4	*P*2$_1$	0*k*0 A_2 B_2 2$_1$ B_3
5	*C*2	
6	*Pm*	
7	*Pc*	*h*0*l*$_o$ A_2 B_2 *c* A_3
8	*Cm*	
9	*Cc*	*h*$_e$0*l*$_o$ A_2 B_2 *c* A_3
10	*P*2/*m*	
11	*P*2$_1$/*m*	0*k*0 A_2 B_2 2$_1$ B_3
12	*C*2/*m*	
13	*P*2/*c*	*h*0*l*$_o$ A_2 B_2 *c* A_3
14	*P*2$_1$/*c*	0*k*0 A_2 B_2 2$_1$ B_3
		*h*0*l*$_o$ A_2 B_2 *c* A_3
15	*C*2/*c*	*h*$_e$0*l*$_o$ A_2 B_2 *c* A_3

Point group 222

Space group		Incident-beam direction [100]	[010]	[001]	[*hk*0]	[0*kl*]	[*h*0*l*]
16	*P*222						
17	*P*222$_1$	00*l* A_2 B_2 2$_1$ B_3	00*l* A_2 B_2 2$_1$ B_3		00*l* A_2 B_2 2$_1$ B_3		
18	*P*2$_1$2$_1$2	0*k*0 A_2 B_2 2$_{12}$ B_3	*h*00 A_2 B_2 2$_{11}$ B_3	*h*00 A_2 B_2 2$_{11}$ B_3 0*k*0 2$_{12}$		*h*00 A_2 B_2 2$_{11}$ B_3	0*k*0 A_2 B_2 2$_{12}$ B_3
19	*P*2$_1$2$_1$2$_1$	0*k*0 A_2 B_2 2$_{12}$ B_3 00*l* 2$_{13}$	*h*00 A_2 B_2 2$_{11}$ B_3 00*l* 2$_{13}$	*h*00 A_2 B_2 2$_{11}$ B_3 0*k*0 2$_{12}$	00*l* A_2 B_2 2$_{13}$ B_3	*h*00 A_2 B_2 2$_{11}$ B_3	0*k*0 A_2 B_2 2$_{12}$ B_3
20	*C*222$_1$	00*l* A_2 B_2 2$_1$ B_3	00*l* A_2 B_2 2$_1$ B_3		00*l* A_2 B_2 2$_1$ B_3		
21	*C*222						
22	*F*222						
23	*I*222						
24	*I*2$_1$2$_1$2$_1$						

Point group *mm*2

Space group		Incident-beam direction [100]	[010]	[001]	[*hk*0]	[0*kl*]	[*h*0*l*]
25	*Pmm*2						
26	*Pmc*2$_1$	00*l* A_2 B_2 *c*, 2$_1$ A_3 B_3	00*l* 2$_1$ B_3		00*l* A_2 B_2 2$_1$ B_3		*h*0*l*$_o$ A_2 B_2 *c* A_3
27	*Pcc*2	00*l* *c*$_2$ A_3	00*l* *c*$_1$ A_3			0*kl*$_o$ A_2 B_2 *c*$_1$ A_3	*h*0*l*$_o$ A_2 B_2 *c*$_2$ A_3
28	*Pma*2			*h*00 A_2 B_2 *a* A_3			*h*$_o$0*l* A_2 B_2 *a* A_3
29	*Pca*2$_1$	00*l* 2$_1$ B_3	00*l* A_2 B_2 *c*, 2$_1$ A_3 B_3	*h*00 A_2 B_2 *a* A_3	00*l* A_2 B_2 2$_1$ B_3	0*kl*$_o$ A_2 B_2 *c* A_3	*h*$_o$0*l* A_2 B_2 *a* A_3

Table 2.5.3.9 (*cont.*)

Space group	Incident-beam direction					
	[100]	[010]	[001]	[*hk*0]	[0*kl*]	[*h*0*l*]
30 *Pnc*2	00*l* *c* A_3	00*l* *n* A_3	0*k*0 A_2 B_2 *n* A_3		0*kl*: A_2 B_2 $k + l =$ A_3 $2n + 1$ *n*	$h0l_o$ A_2 B_2 *c* A_3
31 *Pmn*2_1	00*l* A_2 B_2 *n*, 2_1 A_3 B_3	00*l* 2_1 B_3	*h*00 A_2 B_2 *n* A_3	00*l* A_2 B_2 2_1 B_3		*h*0*l*: A_2 B_2 $h + l =$ A_3 $2n + 1$ *n*
32 *Pba*2			*h*00 A_2 B_2 *a* A_3 0*k*0 *b*		$0k_ol$ A_2 B_2 *b* A_3	h_o0l A_2 B_2 *a* A_3
33 *Pna*2_1	00*l* 2_1 B_3	00*l* A_2 B_2 *n*, 2_1 A_3 B_3	*h*00 A_2 B_2 *a* A_3 0*k*0 *n*	00*l* A_2 B_2 2_1 B_3	0*kl*: A_2 B_2 $k + l =$ A_3 $2n + 1$ *n*	h_o0l A_2 B_2 *a* A_3
34 *Pnn*2	00*l* n_2 A_3	00*l* n_1 A_3	*h*00 A_2 B_2 n_2 A_3 0*k*0 n_1		0*kl*: A_2 B_2 $k + l =$ A_3 $2n + 1$ n_1	*h*0*l*: A_2 B_2 $h + l =$ A_3 $2n + 1$ n_2
35 *Cmm*2 *ba*2						
36 *Cmc*2_1 *bn*2_1	00*l* A_2 B_2 *c*, 2_1 A_3 B_3	00*l* 2_1 B_3		00*l* A_2 B_2 2_1 B_3		h_e0l_o A_2 B_2 *c* A_3
37 *Ccc*2 *nn*2	00*l* c_2 A_3	00*l* c_1 A_3			$0k_el_o$ A_2 B_2 c_1 A_3	h_e0l_o A_2 B_2 c_2 A_3
38 *Amm*2 *nc*2_1						
39 *Abm*2 *cc*2_1					$0k_ol_o$ A_2 B_2 *b* A_3	
40 *Ama*2 *nn*2_1			*h*00 A_2 B_2 *a* A_3			h_o0l_e A_2 B_2 *a* A_3
41 *Aba*2 *cn*2_1			*h*00 A_2 B_2 *a* A_3		$0k_ol_o$ A_2 B_2 *b* A_3	h_o0l_e A_2 B_2 *a* A_3
42 *Fmm*2						
43 *Fdd*2 *dd*2_1	00*l*: $l = 4n + 2$ A_3 d_2	00*l*: $l = 4n + 2$ A_3 d_1	*h*00: A_2 B_2 $h = 4n + 2$ A_3 d_2 0*k*0: $k = 4n + 2$ d_1		$0k_el_e$: A_2 B_2 $k_e + l_e =$ A_3 $4n + 2$ d_1	h_e0l_e: A_2 B_2 $h_e + l_e =$ A_3 $4n + 2$ d_2
44 *Imm*2 *nn*2_1						
45 *Iba*2 *cc*2_1					$0k_ol_o$ A_2 B_2 *b* A_3	h_o0l_o A_2 B_2 *a* A_3
46 *Ima*2 *nc*2_1						h_o0l_o A_2 B_2 *a* A_3

Point group *mmm*

Space group	Incident-beam direction					
	[100]	[010]	[001]	[*hk*0]	[0*kl*]	[*h*0*l*]
47 *P*2/*m*2/*m*2/*m*						
48 *P*2/*n*2/*n*2/*n*	00*l* n_2 A_3 0*k*0 n_3	00*l* n_1 A_3 *h*00 n_3	0*k*0 n_1 A_3 *h*00 n_2	*hk*0: A_2 B_2 $h + k =$ A_3 $2n + 1$ n_3	0*kl*: A_2 B_2 $k + l =$ A_3 $2n + 1$ n_1	*h*0*l*: A_2 B_2 $h + l =$ A_3 $2n + 1$ n_2
49 *P*2/*c*2/*c*2/*m*	00*l* c_2 A_3	00*l* c_1 A_3			$0kl_o$ A_2 B_2 c_1 A_3	$h0l_o$ A_2 B_2 c_2 A_3
50 *P*2/*b*2/*a*2/*n*	0*k*0 *n* A_3	*h*00 *n* A_3	0*k*0 *b* A_3 *h*00 *a*	*hk*0: A_2 B_2 $h + k =$ A_3 $2n + 1$ *n*	$0k_ol$ A_2 B_2 *b* A_3	h_o0l A_2 B_2 *a* A_3
51 $P2_1$/*m*2/*m*2/*a*		*h*00 A_2 B_2 2_1, *a* A_3 B_3	*h*00 2_1 B_3	h_ok0 A_2 B_2 *a* A_3	*h*00 A_2 B_2 2_1 B_3	

Table 2.5.3.9 (*cont.*)

Space group	Incident-beam direction					
	[100]	[010]	[001]	[*hk*0]	[0*kl*]	[*h*0*l*]
52 *P*2/*n*2$_1$/*n*2/*a*	00*l* n_2 A_3	00*l* n_1 A_3 *h*00 *a*	0*k*0 n_1, 2$_1$ A_2 A_3 B_2 B_3	$h_o k0$ *a* A_2 A_3 B_2	0*kl*: $k + l =$ $2n + 1$ n_1 A_2 A_3 B_2	*h*0*l*: $h + l =$ $2n + 1$ n_2 A_2 A_3 B_2
	0*k*0 2$_1$ B_3		*h*00 n_2 A_3			0*k*0 2$_1$ A_2 B_2 B_3
53 *P*2/*m*2/*n*2$_1$/*a*	00*l* n, 2$_1$ A_2 A_3 B_2 B_3	*h*00 *a* A_3	*h*00 *n* A_3	$h_o k0$ *a* A_2 A_3 B_2		*h*0*l*: $h + l =$ $2n + 1$ n A_2 A_3 B_2
		00*l* 2$_1$ B_3		00*l* 2$_1$ A_2 B_2 B_3		
54 *P*2$_1$/*c*2/*c*2/*a*	00*l* c_2 A_3	00*l* c_1 A_3	*h*00 2$_1$ B_3	$h_o k0$ *a* A_2 A_3 B_2	0*kl*$_o$ c_1 A_2 A_3 B_2	*h*0*l*$_o$ c_2 A_2 A_3 B_2
		*h*00 *a*, 2$_1$ A_2 A_3 B_2 B_3			*h*00 2$_1$ A_2 B_2 B_3	
55 *P*2$_1$/*b*2$_1$/*a*2/*m*	0*k*0 2$_{12}$ B_3	*h*00 2$_{11}$ B_3	0*k*0 *b*, 2$_{12}$ A_2 A_3 B_2 B_3		0*k*$_o$*l* *b* A_2 A_3 B_2	*h*$_o$0*l* *a* A_2 A_3 B_2
			*h*00 *a*, 2$_{11}$		*h*00 2$_{11}$ A_2 B_2 B_3	0*k*0 2$_{12}$ A_2 B_2 B_3
56 *P*2$_1$/*c*2$_1$/*c*2/*n*	00*l* c_2 A_3	00*l* c_1 A_3	0*k*0 2$_{12}$ B_3 *h*00 2$_{11}$	*hk*0: $h + k =$ $2n + 1$ n A_2 A_3 B_2	0*kl*$_o$ c_1 A_2 A_3 B_2	*h*0*l*$_o$ c_2 A_2 A_3 B_2
	0*k*0 2$_{12}$, n A_2 A_3 B_2 B_3	*h*00 2$_{11}$, n A_2 A_3 B_2 B_3			*h*00 2$_{11}$ A_2 B_2 B_3	0*k*0 2$_{12}$ A_2 B_2 B_3
57 *P*2/*b*2$_1$/*c*2$_1$/*m*	00*l* c, 2$_{12}$ A_2 A_3 B_2 B_3	00*l* 2$_{12}$ B_3	0*k*0 *b*, 2$_{11}$ A_2 A_3 B_2 B_3	00*l* 2$_{12}$ A_2 B_2 B_3	0*k*$_o$*l* *b* A_2 A_3 B_2	*h*0*l*$_o$ c A_2 A_3 B_2
	0*k*0 2$_{11}$ B_3					0*k*0 2$_{11}$ A_2 B_2 B_3
58 *P*2$_1$/*n*2$_1$/*n*2/*m*	00*l* n_2 A_3	00*l* n_1 A_3	0*k*0 n_1, 2$_{12}$ A_2 A_3 B_2 B_3 *h*00 n_2, 2$_{11}$		0*kl*: $k + l =$ $2n + 1$ n_1 A_2 A_3 B_2	*h*0*l*: $h + l =$ $2n + 1$ n_2 A_2 A_3 B_2
	0*k*0 2$_{12}$ B_3	*h*00 2$_{11}$ B_3			*h*00 2$_{11}$ A_2 B_2 B_3	0*k*0 2$_{12}$ A_2 B_2 B_3
59 *P*2$_1$/*m*2$_1$/*m*2/*n*	0*k*0 n, 2$_{12}$ A_2 A_3 B_2 B_3	*h*00 n, 2$_{11}$ A_2 A_3 B_2 B_3	0*k*0 2$_{12}$ B_3 *h*00 2$_{11}$	*hk*0: $h + k =$ $2n + 1$ n A_2 A_3 B_2	*h*00 2$_{11}$ A_2 B_2 B_3	0*k*0 2$_{12}$ A_2 B_2 A_3
60 *P*2$_1$/*b*2/*c*2$_1$/*n*	00*l* c, 2$_{12}$ A_2 A_3 B_2 B_3	*h*00 n, 2$_{11}$ A_2 A_3 B_2 B_3	0*k*0 *b* A_3	*hk*0: $h + k =$ $2n + 1$ n A_2 A_3 B_2	0*k*$_o$*l* *b* A_2 A_3 B_2	*h*0*l*$_o$ c A_2 A_3 B_2
	0*k*0 n A_3	00*l* 2$_{12}$ B_3	*h*00 2$_{11}$ B_3	00*l* 2$_{12}$ A_2 B_2 B_3	*h*00 2$_{11}$ A_2 B_2 B_3	
61 *P*2$_1$/*b*2$_1$/*c*2$_1$/*a*	00*l* c, 2$_{13}$ A_2 A_3 B_2 B_3	00*l* 2$_{13}$ B_3	0*k*0 *b*, 2$_{12}$ A_2 A_3 B_2 B_3	$h_o k0$ *a* A_2 A_3 B_2	0*k*$_o$*l* *b* A_2 A_3 B_2	*h*0*l*$_o$ c A_2 A_3 B_2
	0*k*0 2$_{12}$ B_3	*h*00 *a*, 2$_{11}$ A_2 A_3 B_2 B_3	*h*00 2$_{11}$ B_3	00*l* 2$_{13}$ A_2 B_2 B_3	*h*00 2$_{11}$ A_2 B_2 B_3	0*k*0 2$_{12}$ A_2 B_2 B_3
62 *P*2$_1$/*n*2$_1$/*m*2$_1$/*a*	00*l* 2$_{13}$ B_3 0*k*0 2$_{12}$	00*l* n, 2$_{13}$ A_2 A_3 B_2 B_3 *h*00 *a*, 2$_{11}$	0*k*0 n, 2$_{12}$ A_2 A_3 B_2 B_3	$h_o k0$ *a* A_2 A_3 B_2	0*kl*: $k + l =$ $2n + 1$ n A_2 A_3 B_2	0*k*0 2$_{12}$ A_2 B_2 B_3
			*h*00 2$_{11}$ B_3	00*l* 2$_{13}$ A_2 B_2 B_3	*h*00 2$_{11}$ A_2 B_2 B_3	
63 *C*2/*m*2/*c*2$_1$/*m*	00*l* c, 2$_1$ A_2 A_3 B_2 B_3	00*l* 2$_1$ B_3		00*l* 2$_1$ A_2 B_2 B_3		*h*$_e$0*l*$_o$ c A_2 A_3 B_2
64 *C*2/*m*2/*c*2$_1$/*a*	00*l* c, 2$_1$ A_2 A_3 B_2 B_3	00*l* 2$_1$ B_3		$h_o k_o 0$ *a* A_2 A_3 B_2		*h*$_e$0*l*$_o$ c A_2 A_3 B_2
				00*l* 2$_1$ A_2 B_2 B_3		
65 *C*2/*m*2/*m*2/*m*						
66 *C*2/*c*2/*c*2/*m*	00*l* c_2 A_3	00*l* c_1 A_3			0*k*$_e$*l*$_o$ c_1 A_2 A_3 B_2	*h*$_e$0*l*$_o$ c_2 A_2 A_3 B_2
67 *C*2/*m*2/*m*2/*a*				$h_o k_o 0$ *a* A_2 A_3 B_2		

Table 2.5.3.9 (*cont.*)

Space group	Incident-beam direction					
	[100]	[010]	[001]	[*hk*0]	[0*kl*]	[*h*0*l*]
68 *C*2/*c*2/*c*2/*a*	00*l* c_2 A_3	00*l* c_1 A_3		$h_o k_o 0$ A_2 B_2 *a* A_3	$0k_e l_o$ A_2 B_2 c_1 A_3	$h_e 0 l_o$ A_2 B_2 c_2 A_3
69 *F*2/*m*2/*m*2/*m*						
70 *F*2/*d*2/*d*2/*d*	00*l*: $l =$ A_3 $4n + 2$ d_2 0*k*0: $k =$ $4n + 2$ d_3	*h*00: $h =$ A_3 $4n + 2$ d_3 00*l*: $l =$ $4n + 2$ d_1	0*k*0: $k =$ A_3 $4n + 2$ d_1 *h*00: $h =$ $4n + 2$ d_2	$h_e k_e 0$: A_2 B_2 $h_e + k_e =$ A_3 $4n + 2$ d_3	$0k_e l_e$: A_2 B_2 $k_e + l_e =$ A_3 $4n + 2$ d_1	$h_e 0 l_e$: A_2 B_2 $h_e + l_e =$ A_3 $4n + 2$ d_2
71 *I*2/*m*2/*m*2/*m*						
72 *I*2/*b*2/*a*2/*m*					$0k_o l_o$ A_2 B_2 *b* A_3	$h_o 0 l_o$ A_2 B_2 *a* A_3
73 $I2_1/b2_1/c2_1/a$				$h_o k_o 0$ A_2 B_2 *a* A_3	$0k_o l_o$ A_2 B_2 *b* A_3	$h_o 0 l_o$ A_2 B_2 *c* A_3
74 $I2_1/m2_1/m2_1/a$				$h_o k_o 0$ A_2 B_2 *a* A_3		

Point groups 4, $\bar{4}$, 4/*m*

Space group	Incident-beam direction
	[*hk*0]
75 *P*4	
76 $P4_1$	00*l* A_2 B_2 4_1 B_3
77 $P4_2$	
78 $P4_3$	00*l* A_2 B_2 4_3 B_3
79 *I*4	
80 $I4_1$	
81 $P\bar{4}$	
82 $I\bar{4}$	
83 *P*4/*m*	
84 $P4_2/m$	
85 *P*4/*n*	*hk*0: $h + k = 2n + 1$ A_2 B_2 *n* A_3
86 $P4_2/n$	*hk*0: $h + k = 2n + 1$ A_2 B_2 *n* A_3
87 *I*4/*m*	
88 $I4_1/a$	$h_o k_o 0$ A_2 B_2 *a* A_3

Point group 422

Space group	Incident-beam direction	
	[*hk*0]	[0*kl*]
89 *P*422		
90 $P42_12$		*h*00 A_2 B_2 2_1 B_3
91 $P4_122$	00*l* A_2 B_2 4_1 B_3	
92 $P4_12_12$	00*l* A_2 B_2 4_1 B_3	*h*00 A_2 B_2 2_1 B_3
93 $P4_222$		
94 $P4_22_12$		*h*00 A_2 B_2 2_1 B_3
95 $P4_322$	00*l* A_2 B_2 4_3 B_3	
96 $P4_32_12$	00*l* A_2 B_2 4_3 B_3	*h*00 A_2 B_2 2_1 B_3
97 *I*422		
98 $I4_122$		

Table 2.5.3.9 (*cont.*)

Point group 4*mm*. The symbol *a* in the column [*h*0*l*] is equivalent to the symbol *b* in the space groups of the first column.

Space group		Incident-beam direction									
		[100]		[001]		[110]		[*h*0*l*]		[*hhl*]	
99	*P*4*mm*										
100	*P*4*bm*			*h*00 a_2 0*k*0 b_1	A_2 B_2 A_3			h_o0l *a*	A_2 B_2 A_3		
101	$P4_2cm$	00*l* c_2 A_3						*h*0l_o *c*	A_2 B_2 A_3		
102	$P4_2nm$	00*l* n_2 A_3		*h*00 n_2 0*k*0 n_1	A_2 B_2 A_3			*h*0*l*: $h + l = 2n + 1$ *n*	A_2 B_2 A_3		
103	*P*4*cc*	00*l* c_{12} A_3				00*l* c_2 A_3		*h*0l_o c_1	A_2 B_2 A_3	*hhl*$_o$ c_2	A_2 B_2 A_3
104	*P*4*nc*	00*l* n_2 A_3		*h*00 n_2 0*k*0 n_1	A_2 B_2 A_3	00*l* *c* A_3		*h*0*l*: $h + l = 2n + 1$ *n*	A_2 B_2 A_3	*hhl*$_o$ *c*	A_2 B_2 A_3
105	$P4_2mc$					00*l* *c* A_3				*hhl*$_o$ *c*	A_2 B_2 A_3
106	$P4_2bc$			*h*00 a_2 0*k*0 b_1	A_2 B_2 A_3	00*l* *c* A_3		h_o0l *a*	A_2 B_2 A_3	*hhl*$_o$ *c*	A_2 B_2 A_3
107	*I*4*mm*										
108	*I*4*cm*							h_o0l_o *c*	A_2 B_2 A_3		
109	$I4_1md$			*hh*0, $\bar{h}h0$ *d*	A_2 B_2 A_3	00*l*: $l = 4n + 2$ *d* A_3				*hhl*$_e$: $2h + l_e = 4n + 2$ *d*	A_2 B_2 A_3
110	$I4_1cd$			*hh*0, $\bar{h}h0$ *d*	A_2 B_2 A_3	00*l*: $l = 4n + 2$ *d* A_3		h_o0l_o *c*	A_2 B_2 A_3	*hhl*$_e$: $2h + l_e = 4n + 2$ *d*	A_2 B_2 A_3

Point group $\bar{4}2m$. The symbol *a* in the column [*h*0*l*] is equivalent to the symbol *b* in the space groups of the first column.

Space group		Incident-beam direction									
		[100]		[001]		[110]		[*h*0*l*]		[*hhl*]	
111	$P\bar{4}2m$										
112	$P\bar{4}2c$					00*l* *c* A_3				*hhl*$_o$ *c*	A_2 B_2 A_3
113	$P\bar{4}2_1m$	0*k*0 A_2 B_2 2_{12} B_3		*h*00 2_{11} 0*k*0 2_{12}	A_2 B_2 B_3			0*k*0 2_1	A_2 B_2 B_3		
114	$P\bar{4}2_1c$	0*k*0 A_2 B_2 2_{12} B_3		*h*00 2_{11} 0*k*0 2_{12}	A_2 B_2 B_3	00*l* *c* A_3		0*k*0 2_1	A_2 B_2 B_3	*hhl*$_o$ *c*	A_2 B_2 A_3
115	$P\bar{4}m2$										
116	$P\bar{4}c2$	00*l* c_2 A_3						*h*0l_o *c*	A_2 B_2 A_3		
117	$P\bar{4}b2$			*h*00 a_2 0*k*0 b_1	A_2 B_2 A_3			h_o0l *a*	A_2 B_2 A_3		
118	$P\bar{4}n2$	00*l* n_2 A_3		*h*00 n_2 0*k*0 n_1	A_2 B_2 A_3			*h*0*l*: $h + l = 2n + 1$ *n*	A_2 B_2 A_3		
119	$I\bar{4}m2$										
120	$I\bar{4}c2$							h_o0l_o *c*	A_2 B_2 A_3		
121	$I\bar{4}2m$										
122	$I\bar{4}2d$			*hh*0, $\bar{h}h0$ *d*	A_2 B_2 A_3	00*l*: $l = 4n + 2$ *d* A_3				*hhl*$_e$: $2h + l_e = 4n + 2$ *d*	A_2 B_2 A_3

Table 2.5.3.9 (*cont.*)

Point group 4/*mmm*. The symbol *a* in the column [*h*0*l*] is equivalent to the symbol *b* in the space groups of the first column.

Space group		Incident-beam direction					
		[100]	[001]	[110]	[*h*0*l*]	[*hhl*]	[*hk*0]
123	*P*4/*mmm* *P*4/*m*2/*m*2/*m*						
124	*P*4/*mcc* *P*4/*m*2/*c*2/*c*	00*l* c_{12} A_3		00*l* c_2 A_3	*h*0l_o A_2 B_2 c_1 A_3	*hhl*$_o$ A_2 B_2 c_2 A_3	
125	*P*4/*nbm* *P*4/*n*2/*b*2/*m*	0*k*0 *n* A_3	*h*00 a_2 A_3 0*k*0 b_1		h_o0*l* *a* A_2 B_2 A_3		*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 *n*
126	*P*4/*nnc* *P*4/*n*2/*n*2/*c*	0*k*0 n_1 A_3 00*l* n_{22}	*h*00 n_{22} A_3 0*k*0 n_{21}	00*l* *c* A_3	*h*0*l*: A_2 B_2 *h* + *l* = 2*n* + 1 A_3 n_2	*hhl*$_o$ A_2 B_2 *c* A_3	*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 n_1
127	*P*4/*mbm* *P*4/*m*2$_1$/*b*2/*m*	0*k*0 2_{12} B_3	*h*00 A_2 B_2 a_2, 2_{11} A_3 B_3 0*k*0 b_1, 2_{12}		*h*$_o$0*l* A_2 B_2 *a* A_3 0*k*0 A_2 B_2 2_1 B_3		
128	*P*4/*mnc* *P*4/*m*2$_1$/*n*2/*c*	00*l* n_2 A_3 0*k*0 2_{12} B_3	*h*00 A_2 B_2 n_2, 2_{11} A_3 B_3 0*k*0 n_1, 2_{12}	00*l* *c* A_3	*h*0*l*: A_2 B_2 *h* + *l* = 2*n* + 1 A_3 *n* 0*k*0 A_2 B_2 2_1 B_3	*hhl*$_o$ A_2 B_2 *c* A_3	
129	*P*4/*nmm* *P*4/*n*2$_1$/*m*2/*m*	0*k*0 A_2 B_2 *n*, 2_{12} A_3 B_3	*h*00 2_{11} B_3 0*k*0 2_{12}		0*k*0 A_2 B_2 2_1 B_3		*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 *n*
130	*P*4/*ncc* *P*4/*n*2$_1$/*c*2/*c*	0*k*0 A_2 B_2 *n*, 2_{12} A_3 B_3 00*l* c_{12} A_3	*h*00 2_{11} B_3 0*k*0 2_{12}	00*l* c_2 A_3	*h*0l_o A_2 B_2 c_1 A_3 0*k*0 A_2 B_2 2_1 B_3	*hhl*$_o$ A_2 B_2 c_2 A_3	*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 *n*
131	*P*4$_2$/*mmc* *P*4$_2$/*m*2/*m*2/*c*			00*l* *c* A_3		*hhl*$_o$ A_2 B_2 *c* A_3	
132	*P*4$_2$/*mcm* *P*4$_2$/*m*2/*c*2/*m*	00*l* c_2 A_3			*h*0l_o A_2 B_2 *c* A_3		
133	*P*4$_2$/*nbc* *P*4$_2$/*n*2/*b*2/*c*	0*k*0 *n* A_3	*h*00 a_2 A_3 0*k*0 b_1	00*l* *c* A_3	h_o0*l* *a* A_2 B_2 A_3	*hhl*$_o$ A_2 B_2 *c* A_3	*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 *n*
134	*P*4$_2$/*nnm* *P*4$_2$/*n*2/*n*2/*m*	0*k*0 n_1 A_3 00*l* n_{22}	*h*00 n_{22} A_3 0*k*0 n_{21}		*h*0*l*: A_2 B_2 *h* + *l* = 2*n* + 1 A_3 n_2		*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 n_1
135	*P*4$_2$/*mbc* *P*4$_2$/*m*2$_1$/*b*2/*c*	0*k*0 2_{12} B_3	*h*00 A_2 B_2 a_2, 2_{11} A_3 B_3 0*k*0 b_1, 2_{12}	00*l* *c* A_3	*h*$_o$0*l* A_2 B_2 *a* A_3 0*k*0 A_2 B_2 2_1 B_3	*hhl*$_o$ A_2 B_2 *c* A_3	
136	*P*4$_2$/*mnm* *P*4$_2$/*m*2$_1$/*n*2/*m*	00*l* n_2 A_3 0*k*0 2_{12} B_3	*h*00 A_2 B_2 n_2, 2_{11} A_3 B_3 0*k*0 n_1, 2_{12}		*h*0*l*: A_2 B_2 *h* + *l* = 2*n* + 1 A_3 *n* 0*k*0 A_2 B_2 2_1 B_3		
137	*P*4$_2$/*nmc* *P*4$_2$/*n*2$_1$/*m*2/*c*	0*k*0 A_2 B_2 *n*, 2_{12} A_3 B_3	*h*00 2_{11} B_3 0*k*0 2_{12}	00*l* *c* A_3	0*k*0 A_2 B_2 2_1 B_3	*hhl*$_o$ A_2 B_2 *c* A_3	*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 *n*
138	*P*4$_2$/*ncm* *P*4$_2$/*n*2$_1$/*c*2/*m*	0*k*0 A_2 B_2 *n*, 2_{12} A_3 B_3 00*l* c_2 A_3	*h*00 2_{11} B_3 0*k*0 2_{12}		*h*0l_o A_2 B_2 *c* A_3 0*k*0 A_2 B_2 2_1 B_3		*hk*0: A_2 B_2 *h* + *k* = 2*n* + 1 A_3 *n*
139	*I*4/*mmm* *I*4/*m*2/*m*2/*m*						

Table 2.5.3.9 (*cont.*)

Space group	Incident-beam direction					
	[100]	[001]	[110]	[*h*0*l*]	[*hhl*]	[*hk*0]
140 *I*4/*mcm* *I*4/*m*2/*c*2/*m*				h_o0l_o A_2 B_2 *c* A_3		
141 $I4_1/amd$ $I4_1/a2/m2/d$		*hh*0, $\bar{h}h0$ *d* A_3	00*l*: *l* = A_3 4*n* + 2 *d* $\bar{h}h0$ *a*		*hhl*$_e$: A_2 B_2 2*h* + l_e = A_3 4*n* + 2 *d*	h_ok0 A_2 B_2 *a* A_3
142 $I4_1/acd$ $I4_1/a2/c2/d$		*hh*0, $\bar{h}h0$ *d* A_3	00*l*: *l* = A_3 4*n* + 2 *d* $\bar{h}h0$ *a*	h_o0l_o A_2 B_2 *c* A_3	*hhl*$_e$: A_2 B_2 2*h* + l_e = A_3 4*n* + 2 *d*	h_ok0 A_2 B_2 *a* A_3

Point groups 3, $\bar{3}$, 32, 3*m*, $\bar{3}m$

Space group	Incident-beam direction	
	[11$\bar{2}$0]	[1$\bar{1}$00]
Nos. 143–155: no GM line		
156 *P*3*m*1		
157 *P*31*m*		
158 *P*3*c*1		00*l* A_2 B_2 *c* A_3
159 *P*31*c*	00*l* A_2 B_2 *c* A_3	
160 *R*3*m*		
161 *R*3*c*		00*l*: *l* = 6*n* + 3 A_2 B_2 *c* A_3
162 $P\bar{3}1m$		
163 $P\bar{3}1c$	00*l* A_2 B_2 *c* A_3	
164 $P\bar{3}m1$		
165 $P\bar{3}c1$		00*l* A_2 B_2 *c* A_3
166 $R\bar{3}m$		
167 $R\bar{3}c$		00*l*: *l* = 6*n* + 3 A_2 B_2 *c* A_3

Point groups 6, $\bar{6}$, 6/*m*, 622, 6*mm*, $\bar{6}m2$, 6/*mmm*

Space group	Incident-beam direction	
	[11$\bar{2}$0]	[1$\bar{1}$00]
168 *P*6		
169 $P6_1$	00*l* A_2 B_2 6_1 B_3	00*l* A_2 B_2 6_1 B_3
170 $P6_5$	00*l* A_2 B_2 6_5 B_3	00*l* A_2 B_2 6_5 B_3
171 $P6_2$		
172 $P6_4$		
173 $P6_3$	00*l* A_2 B_2 6_3 B_3	00*l* A_2 B_2 6_3 B_3
174 $P\bar{6}$		
175 *P*6/*m*		
176 $P6_3/m$	00*l* A_2 B_2 6_3 B_3	00*l* A_2 B_2 6_3 B_3

Table 2.5.3.9 (*cont.*)

Space group		Incident-beam direction					
		[11$\bar{2}$0]			[1$\bar{1}$00]		
177	$P622$						
178	$P6_122$	$00l$ 6_1	A_2	B_2 B_3	$00l$ 6_1	A_2	B_2 B_3
179	$P6_522$	$00l$ 6_5	A_2	B_2 B_3	$00l$ 6_5	A_2	B_2 B_3
180	$P6_222$						
181	$P6_422$						
182	$P6_322$	$00l$ 6_3	A_2	B_2 B_3	$00l$ 6_3	A_2	B_2 B_3
183	$P6mm$						
184	$P6cc$	$00l$ c_2	A_3		$00l$ c_1	A_3	
185	$P6_3cm$	$00l$ 6_3		B_3	$00l$ 6_3, c	A_2 A_3	B_2 B_3
186	$P6_3mc$	$00l$ 6_3, c	A_2 A_3	B_2 B_3	$00l$ 6_3		B_3
187	$P\bar{6}m2$						
188	$P\bar{6}c2$				$00l$ c	A_2 A_3	B_2
189	$P\bar{6}2m$						
190	$P\bar{6}2c$	$00l$ c	A_2 A_3	B_2			
191	$P6/mmm$						
192	$P6/mcc$	$00l$ c_2	A_3		$00l$ c_1	A_3	
193	$P6_3/mcm$	$00l$ 6_3		B_3	$00l$ 6_3, c	A_2 A_3	B_2 B_3
194	$P6_3/mmc$	$00l$ 6_3, c	A_2 A_3	B_2 B_3	$00l$ 6_3		B_3

Point groups 23, $m3$

Space group		Incident-beam direction							
		[100] (cyclic)			[110] (cyclic)			[hk0] (cyclic)		
195	$P23$								
196	$F23$								
197	$I23$								
198	$P2_13$	$00l$ 2_{13} $0k0$ 2_{12}	A_2	B_2 B_3	$00l$ 2_{13}	A_2	B_2 B_3	$00l$ 2_1	A_2	B_2 B_3
199	$I2_13$								
200	$Pm3$ $P2/m\bar{3}$								
201	$Pn3$ $P2/n\bar{3}$	$00l$ n_2 $0k0$ n_3	A_3					$\bar{k}h0$ n	A_2 A_3	B_2
202	$Fm3$ $F2/m\bar{3}$									
203	$Fd3$ $F2/d\bar{3}$	$00l$: $l = 4n + 2$ d_2 $0k0$: $k = 4n + 2$ d_3	A_3					$\bar{k}h0$: $h + k = 4n + 2$ d	A_2 A_3	B_2
204	$Im3$ $I2/m\bar{3}$									
205	$Pa3$ $P2_1/a\bar{3}$	$00l$ c_2, 2_{13}	A_2 A_3	B_2 B_3	$00l$ 2_{13}	A_2	B_2 B_3	$00l$ 2_1	A_2	B_2 B_3
		$0k0$ 2_{12}		B_3	$\bar{h}h0$ a_3	A_2 A_3	B_2	$\bar{k}h0$ a	A_2 A_3	B_2
206	$Ia3$ $I2_1/a\bar{3}$				$\bar{h}h0$ a_3	A_2 A_3	B_2	$\bar{k}h0$ a	A_2 A_3	B_2

Table 2.5.3.9 (*cont.*)

Point group 432

Space group		Incident-beam direction [*hk*0] (cyclic)		
207	$P432$			
208	$P4_232$			
209	$F432$			
210	$F4_132$			
211	$I432$			
212	$P4_332$	$00l$ 4_3	A_2	B_2 B_3
213	$P4_132$	$00l$ 4_1	A_2	B_2 B_3
214	$I4_132$			

Point group $\bar{4}3m$

Space group		Incident-beam direction [100] (cyclic)		[110] (cyclic)		[*hhl*] (cyclic)		
215	$P\bar{4}3m$							
216	$F\bar{4}3m$							
217	$I\bar{4}3m$							
218	$P\bar{4}3n$			$00l$ n	 A_3	hhl_o n	A_2 A_3	B_2
219	$F\bar{4}3c$					$h_oh_ol_o$ c	A_2 A_3	B_2
220	$I\bar{4}3d$	$0kk$, $0\bar{k}k$ d	A_2 A_3	$00l$: $l = 4n + 2$ d	 A_3	hhl_e: $2h + l_e = 4n + 2$ d	A_2 A_3	B_2

Point group $m3m$

Space group		Incident-beam direction [100] (cyclic)		[110] (cyclic)		[*hk*0] (cyclic)			[*hhl*] (cyclic)		
221	$Pm3m$ $P4/m\bar{3}2/m$										
222	$Pn3n$ $P4/n\bar{3}2/n$	$00l$ n_{12} $0k0$ n_{13}	 A_3	$00l$ n_2	 A_3	$hk0$: $h + k = 2n + 1$ n_1	A_2 A_3	B_2	hhl_o n_2	A_2 A_3	B_2
223	$Pm3n$ $P4_2/m\bar{3}2/n$			$00l$ n	 A_3				hhl_o n	A_2 A_3	B_2
224	$Pn3m$ $P4_2/n\bar{3}2/m$	$00l$ n_2 $0k0$ n_3	 A_3			$hk0$: $h + k = 2n + 1$ n	A_2 A_3	B_2			
225	$Fm3m$ $F4/m\bar{3}2/m$										
226	$Fm3c$ $F4/m\bar{3}2/c$								$h_oh_ol_o$ c	A_2 A_3	B_2
227	$Fd3m$ $F4_1/d\bar{3}2/m$	$00l$: $l = 4n + 2$ d_2 $0k0$: $k = 4n + 2$ d_3	 A_3			h_ek_e0: $h_e + k_e = 4n + 2$ d	A_2 A_3	B_2			
228	$Fd3c$ $F4_1/d\bar{3}2/c$	$00l$: $l = 4n + 2$ d_2 $0k0$: $k = 4n + 2$ d_3	 A_3			h_ek_e0: $h_e + k_e = 4n + 2$ d	A_2 A_3	B_2	$h_oh_ol_o$ c	A_2 A_3	B_2
229	$Im3m$ $I4/m\bar{3}2/m$										
230	$Ia3d$ $I4_1/a\bar{3}2/d$	$0kk$, $0\bar{k}k$ d	 A_3	$00l$: $l = 4n + 2$ d $\bar{h}h0$ a_3	 A_3	h_ok_o0 a	A_2 A_3	B_2	hhl_e: $2h + l_e = 4n + 2$ d	A_2 A_3	B_2

Table 2.5.3.10. *Space-group sets indistinguishable by dynamical extinction lines*

(1) $P3$, ($P3_1$, $P3_2$)	(2) $P312$, ($P3_112$, $P3_212$)	(3) $P321$, ($P3_121$, $P3_221$)
(4) $P6$, ($P6_2$, $P6_4$)	(5) $P622$, ($P6_222$, $P6_422$)	(6) $P6_3$, ($P6_1$, $P6_5$)
(7) $P6_322$, ($P6_122$, $P6_522$)	(8) $P4$, $P4_2$	(9) ($P4_1$, $P4_3$)
(10) $P4/m$, $P4_2/m$	(11) $P4/n$, $P4_2/n$	(12) $P422$, $P4_222$
(13) $P42_12$, $P4_22_12$	(14) $I4$, $I4_1$	(15) $I422$, $I4_122$
(16) $I23$, $I2_13$	(17) $I222$, $I2_12_12_1$	(18) $P432$, $P4_232$
(19) ($P4_132$, $P4_332$)	(20) $I432$, $I4_132$	(21) $F432$, $F4_132$
(22) ($P4_122$, $P4_322$)	(23) ($P4_12_12$, $P4_32_12$)	

distinguishes the difference in the relative arrangements of twofold rotation axes and 2_1 screw axes along the [111] direction between the two space groups by examining the symmetry of intensity pairs appearing in the overlapping discs of a coherent [111] ZOLZ pattern. Saitoh, Tsuda *et al.* (2001) extended the method to distinguish the other ten indistinguishable space-group pairs. The method can distinguish between a space group which is composed of a principal rotation axis and a twofold rotation axis like $P321$ and a space group which is composed of a principal screw axis and a twofold rotation axis like $P3_121$ (or $P3_221$) by investigating the difference in the relative arrangements of the twofold rotation axis with respect to the principal axis. Table 2.5.3.11 shows the 12 space-group pairs which are distinguishable by applying the coherent CBED method.

The pairs in parentheses form left- and right-handed space groups. Handedness or chirality may occur in space groups that do not possess mirror and/or inversion symmetry. The handedness of space groups is identified in such a way that the senses of two crystal axes are determined with the aid of kinematical structure-factor calculations and the sense of the third axis is determined with the aid of dynamical calculations. This method was used for quartz by Goodman & Secomb (1977) and Goodman & Johnson (1977) and for MnSi by Tanaka *et al.* (1985). We also mention that Taftø & Spence (1982) developed a simple but clever method without computation for determining the absolute polarity of the sphalerite structure utilizing multiple-scattering effects on weak beams, which are almost independent of thickness. Because of the importance of structure in the field of semiconductor science, this method is conveniently used nowadays to determine polarity.

It is worth mentioning that space groups that are indistinguishable by CBED (Table 2.5.3.10) do not appear frequently in real inorganic materials. The crystal data collected by Nowacki (1967) on 5572 different inorganic materials shows that the number of materials belonging to space groups among sets (2), (3), (5), (7) and (11) in Table 2.5.3.10 is more than 15 but the number belonging to space groups among the other sets is less than ten. This implies that the probability of finding indistinguishable space groups is very low.

Table 2.5.3.11. *Space-group sets distinguishable by coherent CBED*

The space-group pairs in parentheses can not be distinguished by coherent CBED but can be distinguished by a handedness test. An asterisk (*) indicates the incidence at which the distinction is carried out by many-beam interference (Saitoh, Tsuda *et al.*, 2001).

Space-group set	Incidence
(2) $P312$, ($P3_112$, $P3_212$)	$[1\bar{1}01]$
(3) $P321$, ($P3_121$, $P3_221$)	$[11\bar{2}3]$
(5) $P622$, ($P6_222$, $P6_422$)	$[11\bar{2}3]$
(7) $P6_322$, ($P6_122$, $P6_522$)	$[11\bar{2}3]$
(12) $P422$, $P4_222$	[321], [211], [112]*
(13) $P42_12$, $P4_22_12$	[211]
(15) $I422$, $I4_122$	[111]
(16) $I23$, $I2_13$	[111]
(17) $I222$, $I2_12_12_1$	[111]
(18) $P432$, $P4_232$	[321], [211]*
(20) $I432$, $I4_132$	[111]
(21) $F432$, $F4_132$	[432]

2.5.3.3.4. Dynamical extinction in HOLZ reflections

Space-group determination as described in the previous sections is carried out using the extinction lines appearing in ZOLZ reflections. Vertical glide planes whose translation vectors are perpendicular to the specimen surface do not cause extinction lines in ZOLZ reflections but cause them in HOLZ reflections. (It is noted that the vertical glide planes with glide translations not parallel to the surface are not the symmetry elements of diperiodic plane figures.) Vertical glide planes whose translation vectors are parallel to the surface cause extinction lines in both ZOLZ and HOLZ reflections. Vertical screw axes are expected to form extinction lines in HOLZ reflections whose vectors are parallel to the screw axes. These reflections, however, cannot be observed by ordinary CBED. Thus, the extinction lines appearing in observable HOLZ reflections are used to identify not screw axes but glide planes. Examination of HOLZ extinction lines together with ZOLZ extinction lines is an efficient way to characterize the glide vectors and determine the space group.

The dynamical extinction lines appearing in HOLZ reflections caused by the glide planes whose glide vectors are not only parallel but also not parallel to the specimen surface were tabulated by Nagasawa (1983) for various incident-beam orientations of all the space groups that have glide planes. The tabulated results appear on pages 214–225 of the book by Tanaka *et al.* (1988). Table 2.5.3.12 shows the results. The meanings of the letters used in the table are explained in Fig. 2.5.3.13. We consider a vertical glide plane with a glide vector perpendicular to the surface as is shown in Fig. 2.5.3.13(*a*). Letter *A* is given for cases in which the Ewald sphere intersects two circled-cross reflections in the first Laue zone as seen in Fig. 2.5.3.13(*b*), where black circles and circled crosses denote allowed reflections and kinematically forbidden but dynamically allowed reflections due to the glide plane, respectively. *A** denotes cases in which the Ewald sphere intersects a circled-cross reflection on one side of the incident beam and a black-circled reflection on the other, as seen in Fig. 2.5.3.13(*c*). This case occurs only in space group $P2_1/a\bar{3}$. A_h denotes cases in which the Ewald sphere intersects a circled-cross reflection on one side but does not intersect on the other, owing to the asymmetric arrangement of reflections with respect to the incident beam.

The first column of Table 2.5.3.12 list the space groups and the following columns show the type of the extinction lines for possible incident-beam directions. In each pair of columns, the left-hand column gives the reflection indices of the extinction line and the symmetry elements causing the extinction and the right-hand column gives the type of extinction. The first suffix 1, 2 or 3 of a glide symbol distinguishes the first, the second or the third glide plane of a space group. The second suffixes 1 and 2, which appear in the tetragonal and cubic systems, distinguish two equivalent glide planes which lie in the *x* and *y* planes. The suffix o of a reflection index implies that the index is odd-order. Figs.

2.5.3.14(*a*) and (*b*) were taken for FeS_2, space group $P2_1/a\bar{3}$, with incident-beam directions of [100] and [110]. Inserts show enlarged HOLZ patterns for ease of viewing. Extinction lines of type *A* are seen in the $h_o k0$ HOLZ reflections in Fig. 2.5.3.14(*a*) due to the *b*-glide plane (equivalent to the *a*-glide plane in the space-group symbol) parallel to the (001) plane. An extinction line *A** is seen in an $h_o k0$ HOLZ reflection in Fig. 2.5.3.14(*b*) due to the same glide plane as that of Fig. 2.5.3.14(*a*). It should be noted that extinction lines in HOLZ reflections are better observed in thinner specimen areas than those suitable for the observation of the extinction lines in ZOLZ reflections, because the profiles of HOLZ reflections are concentrated into small areas of CBED discs in thicker specimens.

In summary, the use of not only ZOLZ, but also HOLZ extinction lines is recommended for space-group determination.

2.5.3.3.5. Symmetry elements observed by CBED

In the above sections, point-group and space-group determination methods were described following the theory of Buxton *et al.* (1976). They assumed that the observable symmetry elements are those of an infinitely extended parallel-sided specimen or of diperiodic plane figures. CBED patterns determine diffraction groups. Crystal point groups are identified by consulting Fig. 2.5.3.4, which gives the relations between diffraction groups and crystal point groups. When the assumption made by Buxton *et al.* (1976) is accepted in a strict sense, CBED symmetry m_2 caused by a twofold rotation axis oblique to the specimen surface, which is not a symmetry element of a diperiodic plane figure, ought not to be observed. However, the symmetry m_2 due to a twofold rotation axis in the $[1\bar{1}0]$ direction of an Si film with [100] surface normal has been clearly observed at [111] electron incidence (Tanaka *et al.*, 1988, p. 33). This indicates that crystal symmetry elements oblique to the specimen surface are observable when the specimen is tilted. An important condition for CBED is that the top and bottom surfaces be parallel over the specimen area illuminated by the incident beam. CBED observes the symmetry elements of a crystal to the extent that the boundary conditions at the specimen surface do not break the symmetries of the CBED patterns. Gjønnes & Gjønnes (1985) reported that the breaking of CBED symmetry due to a surface oblique to the incident beam is practically negligible.

In Section 2.5.3.3 on space-group determination, space-group symmetry elements of crystals which have glide and screw components parallel to the specimen surface were considered to act as space-group symmetry elements of diperiodic plane figures by mitigating the strict application of the assumption of diperiodic plane figures. In fact, vertical glide planes with a glide vector not parallel to the specimen surface, which were dealt in Section 2.5.3.3.4, are not the symmetry elements of diperiodic plane figures. Ishizuka (1982) showed theoretically that a vertical glide plane with a vertical glide vector produces dynamical extinction lines in HOLZ discs if the Laue zones are well separated. Tanaka *et al.* (1988, pp. 214–225) tabulated the extinction lines appearing in HOLZ discs caused by the vertical glide planes whose glide vectors are not only parallel but also not parallel to the specimen surface. Dynamical extinction lines caused by the glide planes with a glide vector not parallel to the surface have been demonstrated using FeS_2 and $MgAl_2O_4$ (Tanaka *et al.*, 1988, pp. 51–61).

Vertical 2_1, 3_1, 3_2, ..., 6_5 screw axes, which are not symmetry elements of diperiodic plane figures, are expected to form dynamical extinction lines in kinematically forbidden reflections that are located in the direction of the screw axes or of the surface normal. The extinction lines, however, are difficult to observe in ordinary CBED. Thus, CBED does not observe all the symmetry elements of the crystal space groups but observes many more symmetry elements than those of the diperiodic plane figures. It is clear now that it makes no sense to construct space groups using actually observable symmetry elements because they do not form a complete set of groups. It is of no importance to give the relation between the 230 space groups of crystals and the 80 space groups of diperiodic plane figures. Buxton *et al.*'s theory, which determines crystal point groups with the help of diperiodic plane figures, is very beautiful and successful. However, it is not correct to state that CBED observes the symmetry elements of the diperiodic plane figures. The use of the groups of diperiodic plane figures should be recognized as a convention for the sake of convenience. As a further example, horizontal screw axes and horizontal glide planes must be located at the middle of a specimen to form symmetry elements of the diperiodic plane figures. However, those screw axes and glide planes which are not located at the middle of a specimen do produce CBED symmetries. Since we now know that CBED does not observe the symmetries of the diperiodic plane figures but observes those of a physical crystalline specimen, we can determine the corresponding infinite crystal symmetries more freely, by using our knowledge of the symmetries of the sample concerned, guided but not restricted by the beautiful theory of Buxton *et al.* (1976).

One point to note, for symmetry determination, is that one has to be aware of spurious symmetries that appear for crystals of certain structure types (Tanaka *et al.*, 1988, pp. 20–32 and 42–45) and destroy the correct determination of the point and space groups. Another point for precise symmetry determination is that one has to be aware of how CBED symmetry is destroyed by a small breakdown of crystal symmetry (Tanaka *et al.*, 1988 pp. 46–47).

2.5.3.3.6. Examples of space-group determination

A simple example of point-group determination has already been given for Si in Section 2.5.3.2.5. In this section, two examples of space-group determination for rutile and samarium selenide are described, in which the point-group determination still accounts for an important part. The examples look to be a little sophisticated but are a good exercise for those who want to acquire experience in CBED space-group determination. The present determination is carried out by assuming the lattice parameters to be known.

Rutile (TiO_2). The space group of rutile is well known to be $P4_2/mnm$. The lattice parameters are $a = b = 0.459$ nm and $c = 0.296$ nm. Fig. 2.5.3.15(*a*) shows a CBED pattern taken with the [001] incidence at an accelerating voltage of 80 kV. Since no fine HOLZ lines appear in all the discs, projection diffraction groups (column VI of Table 2.5.3.3) have to be applied to explain this pattern. The projection (proj.) WP shows symmetry 4*mm*. The projection diffraction group is found to be $4mm1_R$ from Table 2.5.3.3. Thus, possible diffraction groups are $4m_Rm_R$, 4*mm*, 4_Rmm_R and $4mm1_R$. Another CBED pattern at a second crystal orientation needs to be taken because Fig. 2.5.3.15(*a*) shows only projection symmetry. Figs. 2.5.3.15(*b*) and (*c*) show CBED patterns taken with the [101] incidence at an accelerating voltage of 100 kV. In Fig. 2.5.3.15(*b*), which is the central part of Fig. 2.5.3.15(*c*), no HOLZ lines are seen. The symmetries of the projection BP and projection WP are both 2*mm*. The projection diffraction group of the pattern is $2mm1_R$. The WP of Fig. 2.5.3.15*(c)* is seen to have one mirror symmetry *m*. The diffraction groups which satisfy symmetry *m* are *m*, $m1_R$ and 2_Rmm_R. Among these diffraction groups, the diffraction group whose projection becomes $2mm1_R$ is only diffraction group 2_Rmm_R. By consulting Fig. 2.5.3.4, diffraction group 2_Rmm_R obtained from Figs. 2.5.3.15(*b*) and (*c*) and one diffraction group $4mm1_R$ among diffraction groups $4m_Rm_R$, 4*mm*, 4_Rmm_R and $4mm1_R$ obtained from Fig. 2.5.3.15(*a*) commonly satisfy point group 4/*mmm*. Thus, the point group of rutile is determined to be 4/*mmm*.

Fig. 2.5.3.15(*d*) shows an ordinary diffraction pattern taken with the [001] incidence at an accelerating voltage of 80 kV. With the help of the lattice parameters and the camera length, the indices of the reflections are given as shown in the figure. There

Table 2.5.3.12. *Dynamical extinction lines appearing in HOLZ reflections for crystal space groups that have mirror and glide planes*

Point groups m, $2/m$ (second setting, unique axis b)

Space group		Incident-beam direction [$u0w$]	
6	Pm		
7	Pc	$h0l_o$ c	A_h
8	Cm		
9	Cc	h_e0l_o c	A_h
10	$P2/m$		
11	$P2_1/m$		
12	$C2/m$		
13	$P2/c$	$h0l_o$ c	A_h
14	$P2_1/c$	$h0l_o$ c	A_h
15	$C2/c$	h_e0l_o c	A_h

Point group $mm2$

Space group		Incident-beam direction [100]		[010]		[001]		[$0vw$]		[$u0w$]	
25	$Pmm2$										
26	$Pmc2_1$	$h0l_o$ c	A			$h0l_o$ c	A			$h0l_o$ c	A_h
27	$Pcc2$	$h0l_o$ c_2	A	$0kl_o$ c_1	A	$0kl_o$ c_1 $h0l_o$ c_2	A	$0kl_o$ c_1	A_h	$h0l_o$ c_2	A_h
28	$Pma2$	h_o0l a	A			h_o0l a	A			h_o0l a	A_h
29	$Pca2_1$	h_o0l a	A	$0kl_o$ c	A	$0kl_o$ c h_o0l a	A	$0kl_o$ c	A_h	h_o0l a	A_h
30	$Pnc2$	$h0l_o$ c	A	$0kl$: $k + l = 2n + 1$ n	A	$0kl$: $k + l = 2n + 1$ n $h0l_o$ c	A	$0kl$: $k + l = 2n + 1$ n	A_h	$h0l_o$ c	A_h
31	$Pmn2_1$	$h0l$: $h + l = 2n + 1$ n	A			$h0l$: $h + l = 2n + 1$ n	A			$h0l$: $h + l = 2n + 1$ n	A_h
32	$Pba2$	h_o0l a	A	$0k_ol$ b	A	$0k_ol$ b h_o0l a	A	$0k_ol$ b	A_h	h_o0l a	A_h
33	$Pna2_1$	h_o0l a	A	$0kl$: $k + l = 2n + 1$ n	A	$0kl$: $k + l = 2n + 1$ n h_o0l a	A	$0kl$: $k + l = 2n + 1$ n	A_h	h_o0l a	A_h
34	$Pnn2$	$h0l$: $h + l = 2n + 1$ n_2	A	$0kl$: $k + l = 2n + 1$ n_1	A	$0kl$: $k + l = 2n + 1$ n_1 $h0l$: $h + l = 2n + 1$ n_2	A	$0kl$: $k + l = 2n + 1$ n_1	A_h	$h0l$: $h + l = 2n + 1$ n_2	A_h
35	$Cmm2$ $ba2$										
36	$Cmc2_1$ $bn2_1$	h_e0l_o c	A			h_e0l_o c	A			h_e0l_o c	A_h
37	$Ccc2$ $nn2$	h_e0l_o c_2	A	$0k_el_o$ c_1	A	$0k_el_o$ c_1 h_e0l_o c_2	A	$0k_el_o$ c_1	A_h	h_e0l_o c_2	A_h

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction									
		[100]		[010]		[001]		[0*vw*]		[*u*0*w*]	
38	*Amm*2 *nc*2_1										
39	*Abm*2 *cc*2_1			$0k_ol_o$ *b*	*A*	$0k_ol_o$ *b*	*A*	$0k_ol_o$ *b*	A_h		
40	*Ama*2 *nn*2_1	h_o0l_e *a*	*A*			h_o0l_e *a*	*A*			h_o0l_e *a*	A_h
41	*Aba*2 *cn*2_1	h_o0l_e *a*	*A*	$0k_ol_o$ *b*	*A*	$0k_ol_o$ *b* h_o0l_e *a*	*A*	$0k_ol_o$ *b*	A_h	h_o0l_e *a*	A_h
42	*Fmm*2										
43	*Fdd*2 *dd*2_1	h_e0l_e: $h_e + l_e = 4n + 2$ d_2	*A*	$0k_el_e$: $k_e + l_e = 4n + 2$ d_1	*A*	$0k_el_e$: $k_e + l_e = 4n + 2$ d_1 h_e0l_e: $h_e + l_e = 4n + 2$ d_2	*A*	$0k_el_e$: $k_e + l_e = 4n + 2$ d_1	A_h	h_e0l_e: $h_e + l_e = 4n + 2$ d_2	A_h
44	*Imm*2 *nn*2_1										
45	*Iba*2 *cc*2_1	h_o0l_o *a*	*A*	$0k_ol_o$ *b*	*A*	$0k_ol_o$ *b* h_o0l_o *a*	*A*	$0k_ol_o$ *b*	A_h	h_o0l_o *a*	A_h
46	*Ima*2 *nc*2_1	h_o0l_o *a*	*A*			h_o0l_o *a*	*A*			h_o0l_o *a*	A_h

Point group *mmm*

Space group		Incident-beam direction											
		[100]		[010]		[001]		[*uv*0]		[0*vw*]		[*u*0*w*]	
47	*P*2/*m*2/*m*2/*m*												
48	*P*2/*n*2/*n*2/*n*	*h*0*l*: $h + l = 2n + 1$ n_2 *hk*0: $h + k = 2n + 1$ n_3	*A*	0*kl*: $k + l = 2n + 1$ n_1 *hk*0: $h + k = 2n + 1$ n_3	*A*	0*kl*: $k + l = 2n + 1$ n_1 *h*0*l*: $h + l = 2n + 1$ n_2	*A*	*hk*0: $h + k = 2n + 1$ n_3	A_h	0*kl*: $k + l = 2n + 1$ n_1	A_h	*h*0*l*: $h + l = 2n + 1$ n_2	A_h
49	*P*2/*c*2/*c*2/*m*	$h0l_o$ c_2	*A*	$0kl_o$ c_1	*A*	$0kl_o$ c_1 $h0l_o$ c_2	*A*			$0kl_o$ c_1	A_h	$h0l_o$ c_2	A_h
50	*P*2/*b*2/*a*2/*n*	h_o0l *a* *hk*0: $h + k = 2n + 1$ *n*	*A*	$0k_ol$ *b* *hk*0: $h + k = 2n + 1$ *n*	*A*	$0k_ol$ *b* h_o0l *a*	*A*	*hk*0: $h + k = 2n + 1$ *n*	A_h	$0k_ol$ *b*	A_h	h_o0l *a*	A_h
51	$P2_1$/*m*2/*m*2/*a*	h_ok0 *a*	*A*	h_ok0 *a*	*A*			h_ok0 *a*	A_h				
52	*P*2/*n*2_1/*n*2/*a*	*h*0*l*: $h + l = 2n + 1$ n_2 h_ok0 *a*	*A*	0*kl*: $k + l = 2n + 1$ n_1 h_ok0 *a*	*A*	0*kl*: $k + l = 2n + 1$ n_1 *h*0*l*: $h + l = 2n + 1$ n_2	*A*	h_ok0 *a*	A_h	0*kl*: $k + l = 2n + 1$ n_1	A_h	*h*0*l*: $h + l = 2n + 1$ n_2	A_h
53	*P*2/*m*2/*n*2_1/*a*	*h*0*l*: $h + l = 2n + 1$ *n* h_ok0 *a*	*A*	h_ok0 *a*	*A*	*h*0*l*: $h + l = 2n + 1$ *n*	*A*	h_ok0 *a*	A_h			*h*0*l*: $h + l = 2n + 1$ *n*	A_h
54	$P2_1$/*c*2/*c*2/*a*	$h0l_o$ c_2 h_ok0 *a*	*A*	$0kl_o$ c_1 h_ok0 *a*	*A*	$0kl_o$ c_1 $h0l_o$ c_2	*A*	h_ok0 *a*	A_h	$0kl_o$ c_1	A_h	$h0l_o$ c_2	A_h

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction											
		[100]		[010]		[001]		[uv0]		[0vw]		[u0w]	
55	$P2_1/b2_1/a2/m$	h_o0l a	A	$0k_ol$ b	A	$0k_ol$ b h_o0l a	A			$0k_ol$ b	A_h	h_o0l a	A_h
56	$P2_1/c2_1/c2/n$	$h0l_o$ c_2 $hk0$: $h + k =$ $2n + 1$ n	A	$0kl_o$ c_1 $hk0$: $h + k =$ $2n + 1$ n	A	$0kl_o$ c_1 $h0l_o$ c_2	A	$hk0$: $h + k =$ $2n + 1$ n	A_h	$0kl_o$ c_1	A_h	$h0l_o$ c_2	A_h
57	$P2/b2_1/c2_1/m$	$h0l_o$ c	A	$0k_ol$ b	A	$0k_ol$ b $h0l_o$ c	A			$0k_ol$ b	A_h	$h0l_o$ c	A_h
58	$P2_1/n2_1/n2/m$	$h0l$: $h + l =$ $2n + 1$ n_2	A	$0kl$: $k + l =$ $2n + 1$ n_1	A	$0kl$: $k + l =$ $2n + 1$ n_1 $h0l$: $h + l =$ $2n + 1$ n_2	A			$0kl$: $k + l =$ $2n + 1$ n_1	A_h	$h0l$: $h + l =$ $2n + 1$ n_2	A_h
59	$P2_1/m2_1/m2/n$	$hk0$: $h + k =$ $2n + 1$ n	A	$hk0$: $h + k =$ $2n + 1$ n	A			$hk0$: $h + k =$ $2n + 1$ n	A_h				
60	$P2_1/b2/c2_1/n$	$h0l_o$ c $hk0$: $h + k =$ $2n + 1$ n	A	$0k_ol$ b $hk0$: $h + k =$ $2n + 1$ n	A	$0k_ol$ b $h0l_o$ c	A	$hk0$: $h + k =$ $2n + 1$ n	A_h	$0k_ol$ b	A_h	$h0l_o$ c	A_h
61	$P2_1/b2_1/c2_1/a$	$h0l_o$ c h_ok0 a	A	$0k_ol$ b h_ok0 a	A	$0k_ol$ b $h0l_o$ c	A	h_ok0 a	A_h	$0k_ol$ b	A_h	$h0l_o$ c	A_h
62	$P2_1/n2_1/m2_1/a$	h_ok0 a	A	$0kl$: $k + l =$ $2n + 1$ n h_ok0 a	A	$0kl$: $k + l =$ $2n + 1$ n	A	h_ok0 a	A_h	$0kl$: $k + l =$ $2n + 1$ n	A_h		
63	$C2/m2/c2_1/m$	h_e0l_o c	A			h_e0l_o c	A					h_e0l_o c	A_h
64	$C2/m2/c2_1/a$	h_e0l_o c h_ok_o0 a	A	h_ok_o0 a	A	h_e0l_o c	A	h_ok_o0 a	A_h			h_e0l_o c	A_h
65	$C2/m2/m2/m$												
66	$C2/c2/c2/m$	h_e0l_o c_2	A	$0k_el_o$ c_1	A	$0k_el_o$ c_1 h_e0l_o c_2	A			$0k_el_o$ c_1	A_h	h_e0l_o c_2	A_h
67	$C2/m2/m2/a$	h_ok_o0 a	A	h_ok_o0 a	A			h_ok_o0 a	A_h				
68	$C2/c2/c2/a$	h_e0l_o c_2 h_ok_o0 a	A	$0k_el_o$ c_1 h_ok_o0 a	A	$0k_el_o$ c_1 h_e0l_o c_2	A	h_ok_o0 a	A_h	$0k_el_o$ c_1	A_h	h_e0l_o c_2	A_h
69	$F2/m2/m2/m$												
70	$F2/d2/d2/d$	h_e0l_e: $h_e + l_e =$ $4n + 2$ d_2 h_ek_e0: $h_e + k_e =$ $4n + 2$ d_3	A	h_ek_e0: $h_e + k_e =$ $4n + 2$ d_3 $0k_el_e$: $k_e + l_e =$ $4n + 2$ d_1	A	$0k_el_e$: $k_e + l_e =$ $4n + 2$ d_1 h_e0l_e: $h_e + l_e =$ $4n + 2$ d_2	A	h_ek_e0: $h_e + k_e =$ $4n + 2$ d_3	A_h	$0k_el_e$: $k_e + l_e =$ $4n + 2$ d_1	A_h	h_e0l_e: $h_e + l_e =$ $4n + 2$ d_2	A_h
71	$I2/m2/m2/m$												

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction											
		[100]		[010]		[001]		[uv0]		[0vw]		[u0w]	
72	$I2/b2/a2/m$	h_o0l_o a	A	$0k_ol_o$ b	A	$0k_ol_o$ b h_o0l_o a	A			$0k_ol_o$ b	A_h	h_o0l_o a	A_h
73	$I2_1/b2_1/c2_1/a$	h_o0l_o c h_ok_o0 a	A	h_ok_o0 a $0k_ol_o$ b	A	$0k_ol_o$ b h_o0l_o c	A	h_ok_o0 a	A_h	$0k_ol_o$ b	A_h	h_o0l_o c	A_h
74	$I2_1/m2_1/m2_1/a$	h_ok_o0 a	A	h_ok_o0 a	A			h_ok_o0 a	A_h				

Point group 4/m

Space group		Incident-beam direction			
		[100], [110]		[uv0]	
83	$P4/m$				
84	$P4_2/m$				
85	$P4/n$	$hk0$: $h + k = 2n + 1$ n	A	$hk0$: $h + k = 2n + 1$ n	A_h
86	$P4_2/n$	$hk0$: $h + k = 2n + 1$ n	A	$hk0$: $h + k = 2n + 1$ n	A_h
87	$I4/m$				
88	$I4_1/a$	h_ok_o0 a	A	h_ok_o0 a	A_h

Point group 4mm. The symbol a in the column [u0w] is equivalent to the symbol b in the space groups of the first column.

Space group		Incident-beam direction									
		[100]		[001]		[110]		[u0w]		[uuw]	
99	$P4mm$										
100	$P4bm$	h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2	A			h_o0l a	A_h		
101	$P4_2cm$	$h0l_o$ c_2	A	$0kl_o$ c_1 $h0l_o$ c_2	A			$h0l_o$ c	A_h		
102	$P4_2nm$	$h0l$: $h + l = 2n + 1$ n_2	A	$0kl$: $k + l = 2n + 1$ n_1 $h0l$: $h + l = 2n + 1$ n_2	A			$h0l$: $h + l = 2n + 1$ n	A_h		
103	$P4cc$	$h0l_o$ c_{12}	A	$0kl_o$ c_{11} $h0l_o$ c_{12} $hhl_o, \bar{h}hl_o$ c_2	A	hhl_o c_2	A	$h0l_o$ c_1	A_h	hhl_o c_2	A_h
104	$P4nc$	$h0l$: $h + l = 2n + 1$ n_2	A	$0kl$: $k + l = 2n + 1$ n_1 $h0l$: $h + l = 2n + 1$ n_2 $hhl_o, \bar{h}hl_o$ c	A	hhl_o c	A	$h0l$: $h + l = 2n + 1$ n	A_h	hhl_o c	A_h
105	$P4_2mc$			$hhl_o, \bar{h}hl_o$ c	A	hhl_o c	A			hhl_o c	A_h
106	$P4_2bc$	h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2 $hhl_o, \bar{h}hl_o$ c	A	hhl_o c	A	h_o0l a	A_h	hhl_o c	A_h
107	$I4mm$										

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction									
		[100]		[001]		[110]		[$u0w$]		[uuw]	
108	$I4cm$	h_o0l_o c_2	A	$0k_ol_o$ c_1 h_o0l_o c_2	A			h_o0l_o c	A_h		
109	$I4_1md$			hhl_e, $\bar{h}hl_e$: $2h + l_e = 4n + 2$ d	A	hhl_e: $2h + l_e = 4n + 2$ d	A			hhl_e: $2h + l_e = 4n + 2$ d	A_h
110	$I4_1cd$	h_o0l_o c_2	A	$0k_ol_o$ c_1 h_o0l_o c_2 hhl_e, $\bar{h}hl_e$: $2h + l_e = 4n + 2$ d	A	hhl_e: $2h + l_e = 4n + 2$ d	A	h_o0l_o c	A_h	hhl_e: $2h + l_e = 4n + 2$ d	A_h

Point group $\bar{4}2m$. The symbol a in the column [$u0w$] is equivalent to the symbol b in the space groups of the first column.

Space group		Incident-beam direction									
		[100]		[001]		[110]		[$u0w$]		[uuw]	
111	$P\bar{4}2m$										
112	$P\bar{4}2c$			hhl_o, $\bar{h}hl_o$ c	A	hhl_o c	A			hhl_o c	A_h
113	$P\bar{4}2_1m$										
114	$P\bar{4}2_1c$			hhl_o, $\bar{h}hl_o$ c	A	hhl_o c	A			hhl_o c	A_h
115	$P\bar{4}m2$										
116	$P\bar{4}c2$	$h0l_o$ c_2	A	$0kl_o$ c_1 $h0l_o$ c_2	A			$h0l_o$ c	A_h		
117	$P\bar{4}b2$	h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2	A			h_o0l a	A_h		
118	$P\bar{4}n2$	$h0l$: $h + l = 2n + 1$ n_2	A	$0kl$: $k + l = 2n + 1$ n_1 $h0l$: $h + l = 2n + 1$ n_2	A			$h0l$: $h + l = 2n + 1$ n	A_h		
119	$I\bar{4}m2$										
120	$I\bar{4}c2$	h_o0l_o c_2	A	$0k_ol_o$ c_1 h_o0l_o c_2	A			h_o0l_o c	A_h		
121	$I\bar{4}2m$										
122	$I\bar{4}2d$			hhl_e, $\bar{h}hl_e$: $2h + l_e = 4n + 2$ d	A	hhl_e: $2h + l_e = 4n + 2$ d	A			hhl_e: $2h + l_e = 4n + 2$ d	A_h

Point group $4/mmm$. The symbol a in the column [$u0w$] is equivalent to the symbol b in the space groups of the first column.

Space group		Incident-beam direction											
		[100]		[001]		[110]		[$u0w$]		[uuw]		[$uv0$]	
123	$P4/mmm$ $P4/m2/m2/m$												
124	$P4/mcc$ $P4/m2/c2/c$	$h0l_o$ c_{12}	A	$0kl_o$ c_{11} $h0l_o$ c_{12} hhl_o, $\bar{h}hl_o$ c_2	A	hhl_o c_2	A	$h0l_o$ c_1	A_h	hhl_o c_2	A_h		
125	$P4/nbm$ $P4/n2/b2/m$	$hk0$: $h + k =$ $2n + 1$ n h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2	A	$hk0$: $h + k =$ $2n + 1$ n	A	h_o0l a	A_h			$hk0$: $h + k =$ $2n + 1$ n	A_h

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction											
		[100]		[001]		[110]		[u0w]		[uuw]		[uv0]	
126	*P*4/*nnc* *P*4/*n*2/*n*2/*c*	$hk0$: $h + k =$ $2n + 1$ n_1 $h0l$: $h + l =$ $2n + 1$ n_{22}	A	$0kl$: $k + l =$ $2n + 1$ n_{21} $h0l$: $h + l =$ $2n + 1$ n_{22} $hhl_o, \bar{h}hl_o$ c	A	$hk0$: $h + k =$ $2n + 1$ n_1 hhl_o c	A	$h0l$: $h + l =$ $2n + 1$ n_2	A_h	hhl_o c	A_h	$hk0$: $h + k =$ $2n + 1$ n_1	A_h
127	*P*4/*mbm* *P*4/*m*2$_1$/*b*2/*m*	h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2	A			h_o0l a	A_h				
128	*P*4/*mnc* *P*4/*m*2$_1$/*n*2/*c*	$h0l$: $h + l =$ $2n + 1$ n_2	A	$0kl$: $k + l =$ $2n + 1$ n_1 $h0l$: $h + l =$ $2n + 1$ n_2 $hhl_o, \bar{h}hl_o$ c	A	hhl_o c	A	$h0l$: $h + l =$ $2n + 1$ n	A_h	hhl_o c	A_h		
129	*P*4/*nmm* *P*4/*n*2$_1$/*m*2/*m*	$hk0$: $h + k =$ $2n + 1$ n	A			$hk0$: $h + k =$ $2n + 1$ n	A					$hk0$: $h + k =$ $2n + 1$ n	A_h
130	*P*4/*ncc* *P*4/*n*2$_1$/*c*2/*c*	$hk0$: $h + k =$ $2n + 1$ n $h0l_o$ c_{12}	A	$0kl_o$ c_{11} $h0l_o$ c_{12} $hhl_o, \bar{h}hl_o$ c_2	A	$hk0$: $h + k =$ $2n + 1$ n hhl_o c_2	A	$h0l_o$ c_1	A_h	hhl_o c_2	A_h	$hk0$: $h + k =$ $2n + 1$ n	A_h
131	*P*4$_2$/*mmc* *P*4$_2$/*m*2/*m*2/*c*			$hhl_o, \bar{h}hl_o$ c	A	hhl_o c	A			hhl_o c	A_h		
132	*P*4$_2$/*mcm* *P*4$_2$/*m*2/*c*2/*m*	$h0l_o$ c_2	A	$0kl_o$ c_1 $h0l_o$ c_2	A			$h0l_o$ c	A_h				
133	*P*4$_2$/*nbc* *P*4$_2$/*n*2/*b*2/*c*	$hk0$: $h + k =$ $2n + 1$ n h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2 $hhl_o, \bar{h}hl_o$ c	A	$hk0$: $h + k =$ $2n + 1$ n hhl_o c	A	h_o0l a	A_h	hhl_o c	A_h	$hk0$: $h + k =$ $2n + 1$ n	A_h
134	*P*4$_2$/*nnm* *P*4$_2$/*n*2/*n*2/*m*	$hk0$: $h + k =$ $2n + 1$ n_1 $h0l$: $h + l =$ $2n + 1$ n_{22}	A	$0kl$: $k + l =$ $2n + 1$ n_{21} $h0l$: $h + l =$ $2n + 1$ n_{22}	A	$hk0$: $h + k =$ $2n + 1$ n_1	A	$h0l$: $h + l =$ $2n + 1$ n_2	A_h			$hk0$: $h + k =$ $2n + 1$ n_1	A_h
135	*P*4$_2$/*mbc* *P*4$_2$/*m*2$_1$/*b*2/*c*	h_o0l a_2	A	$0k_ol$ b_1 h_o0l a_2 $hhl_o, \bar{h}hl_o$ c	A	hhl_o c	A	h_o0l a	A_h	hhl_o c	A_h		
136	*P*4$_2$/*mnm* *P*4$_2$/*m*2$_1$/*n*2/*m*	$h0l$: $h + l =$ $2n + 1$ n_2	A	$0kl$: $k + l =$ $2n + 1$ n_1 $h0l$: $h + l =$ $2n + 1$ n_2	A			$h0l$: $h + l =$ $2n + 1$ n	A_h				

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction											
		[100]		[001]		[110]		[*u*0*w*]		[*uuw*]		[*uv*0]	
137	$P4_2/nmc$ $P4_2/n2_1/m2/c$	$hk0$: $h + k = 2n + 1$ n	A	hhl_o, $\bar{h}hl_o$ c	A	hhl_o c $hk0$: $h + k = 2n + 1$ n	A			hhl_o c	A_h	$hk0$: $h + k = 2n + 1$ n	A_h
138	$P4_2/ncm$ $P4_2/n2_1/c2/m$	$hk0$: $h + k = 2n + 1$ n $h0l_o$ c_2	A	$0kl_o$ c_1 $h0l_o$ c_2	A	$hk0$: $h + k = 2n + 1$ n	A	$h0l_o$ c	A_h			$hk0$: $h + k = 2n + 1$ n	A_h
139	$I4/mmm$ $I4/m2/m2/m$												
140	$I4/mcm$ $I4/m2/c2/m$	h_o0l_o c_2	A	$0k_ol_o$ c_1 h_o0l_o c_2	A			h_o0l_o c	A_h				
141	$I4_1/amd$ $I4_1/a2/m2/d$	h_ok_o0 a	A	hhl_e, $\bar{h}hl_e$: $2h + l_e = 4n + 2$ d	A	h_ok_o0 a hhl_e: $2h + l_e = 4n + 2$ d	A			hhl_e: $2h + l_e = 4n + 2$ d	A_h	h_ok_o0 a	A_h
142	$I4_1/acd$ $I4_1/a2/c2/d$	h_ok_o0 a h_o0l_o c_2	A	$0k_ol_o$ c_1 h_o0l_o c_2 hhl_e, $\bar{h}hl_e$: $2h + l_e = 4n + 2$ d	A	h_ok_o0 a hhl_e: $2h + l_e = 4n + 2$ d	A	h_o0l_o c	A_h	hhl_e: $2h + l_e = 4n + 2$ d	A_h	h_ok_o0 a	A_h

Point groups $3m$, $\bar{3}m$

Space group		Incident-beam direction									
		[0001]		[11$\bar{2}$0]		[1$\bar{1}$00]		[11$\bar{2}w$]		[1$\bar{1}0w$]	
156	$P3m1$										
157	$P31m$										
158	$P3c1$	$h\bar{h}0l_o$, $0h\bar{h}l_o$, $\bar{h}0hl_o$ c	A			$h\bar{h}0l_o$ c	A			$h\bar{h}0l_o$ c	A_h
159	$P31c$	$hh\overline{2h}l_o$, $h\overline{2h}hl_o$, $\overline{2h}hhl_o$ c	A	$hh\overline{2h}l_o$ c	A			$hh\overline{2h}l_o$ c	A_h		
160	$R3m$										
161	$R3c$	$h\bar{h}0l_o$, $0h\bar{h}l_o$, $\bar{h}0hl_o$: $h + l_o = 3n$ c	A_h			$h\bar{h}0l_o$: $h + l_o = 3n$ c	A_h			$h\bar{h}0l_o$: $h + l_o = 3n$ c	A_h
162	$P\bar{3}1m$										
163	$P\bar{3}1c$	$hh\overline{2h}l_o$, $h\overline{2h}hl_o$, $\overline{2h}hhl_o$ c	A	$hh\overline{2h}l_o$ c	A			$hh\overline{2h}l_o$ c	A_h		
164	$P\bar{3}m1$										
165	$P\bar{3}c1$	$h\bar{h}0l_o$, $0h\bar{h}l_o$, $\bar{h}0hl_o$ c	A			$h\bar{h}0l_o$ c	A			$h\bar{h}0l_o$ c	A_h
166	$R\bar{3}m$										
167	$R\bar{3}c$	$h\bar{h}0l_o$, $0h\bar{h}l_o$, $\bar{h}0hl_o$: $h + l_o = 3n$ c	A_h			$h\bar{h}0l_o$: $h + l_o = 3n$ c	A_h			$h\bar{h}0l_o$: $h + l_o = 3n$ c	A_h

Point groups $6mm$, $\bar{6}m2$, $6/mmm$

Space group		Incident-beam direction									
		[0001]		[11$\bar{2}$0]		[1$\bar{1}$00]		[11$\bar{2}w$]		[1$\bar{1}0w$]	
183	$P6mm$										
184	$P6cc$	$h\bar{h}0l_o$, $0h\bar{h}l_o$, $\bar{h}0hl_o$ c_1 $hh\overline{2h}l_o$, $h\overline{2h}hl_o$, $\overline{2h}hhl_o$ c_2	A	$hh\overline{2h}l_o$ c_2	A	$h\bar{h}0l_o$ c_1	A	$hh\overline{2h}l_o$ c_2	A_h	$h\bar{h}0l_o$ c_1	A_h

Table 2.5.3.12 (*cont.*)

Space group		Incident-beam direction									
		[0001]		[11$\bar{2}$0]		[1$\bar{1}$00]		[11$\bar{2}$*w*]		[1$\bar{1}$0*w*]	
185	*P*6$_3$*cm*	$hh\bar{2}\bar{h}l_o, h\bar{2}\bar{h}hl_o, \bar{2}\bar{h}hhl_o$ *c*	*A*			$h\bar{h}0l_o$ *c*	*A*			$h\bar{h}0l_o$ *c*	A_h
186	*P*6$_3$*mc*	$hh\bar{2}\bar{h}l_o, h\bar{2}\bar{h}hl_o, \bar{2}\bar{h}hhl_o$ *c*	*A*	$hh\bar{2}\bar{h}l_o$ *c*	*A*			$hh\bar{2}\bar{h}l_o$ *c*	A_h		
187	*P*$\bar{6}$*m*2										
188	*P*$\bar{6}$*c*2	$h\bar{h}0l_o, 0h\bar{h}l_o, \bar{h}0hl_o$ *c*	*A*			$h\bar{h}0l_o$ *c*	*A*			$h\bar{h}0l_o$ *c*	A_h
189	*P*$\bar{6}$2*m*										
190	*P*$\bar{6}$2*c*	$hh\bar{2}\bar{h}l_o, h\bar{2}\bar{h}hl_o, \bar{2}\bar{h}hhl_o$ *c*	*A*	$hh\bar{2}\bar{h}l_o$ *c*	*A*			$hh\bar{2}\bar{h}l_o$ *c*	A_h		
191	*P*6/*mmm*										
192	*P*6/*mcc*	$h\bar{h}0l_o, 0h\bar{h}l_o, \bar{h}0hl_o$ c_1 $hh\bar{2}\bar{h}l_o, h\bar{2}\bar{h}hl_o, \bar{2}\bar{h}hhl_o$ c_2	*A*	$hh\bar{2}\bar{h}l_o$ c_2	*A*	$h\bar{h}0l_o$ c_1	*A*	$hh\bar{2}\bar{h}l_o$ c_2	A_h	$h\bar{h}0l_o$ c_1	A_h
193	*P*6$_3$/*mcm*	$h\bar{h}0l_o, 0h\bar{h}l_o, \bar{h}0hl_o$ *c*	*A*			$h\bar{h}0l_o$ *c*	*A*			$h\bar{h}0l_o$ *c*	A_h
194	*P*6$_3$/*mmc*	$hh\bar{2}\bar{h}l_o, h\bar{2}\bar{h}hl_o, \bar{2}\bar{h}hhl_o$ *c*	*A*	$hh\bar{2}\bar{h}l_o$ *c*	*A*			$hh\bar{2}\bar{h}l_o$ *c*	A_h		

Point group *m*3

Space group		Incident-beam direction					
		[100]		[110]		[*uv*0]	
200	*Pm*3 *P*2/*m*$\bar{3}$						
201	*Pn*3 *P*2/*n*$\bar{3}$	*h*0*l*: $h + l = 2n + 1$ n_2 *hk*0: $h + k = 2n + 1$ n_3	*A*	*hk*0: $h + k = 2n + 1$ n_3	*A*	*hk*0: $h + k = 2n + 1$ *n*	A_h
202	*Fm*3 *F*2/*m*$\bar{3}$						
203	*Fd*3 *F*2/*d*$\bar{3}$	h_e0l_e: $h_e + l_e = 4n + 2$ d_2 h_ek_e0: $h_e + k_e = 4n + 2$ d_3	*A*	h_ek_e0: $h_e + k_e = 4n + 2$ d_3	*A*	h_ek_e0: $h_e + k_e = 4n + 2$ *d*	A_h
204	*Im*3 *I*2/*m*$\bar{3}$						
205	*Pa*3 *P*2$_1$/*a*$\bar{3}$	$h0l_o$ c_2 h_ok0 a_3	*A*	h_ok0 a_3	*A**	h_ok0 *a*	A_h
206	*Ia*3 *I*2$_1$/*a*$\bar{3}$	h_o0l_o c_2 h_ok_o0 a_3	*A*	h_ok_o0 a_3	*A*	h_ok_o0 *a*	A_h

Point group $\bar{4}3m$. The symbol *a* in the column [100] is equivalent to the symbol *c* in the space groups of the first column.

Space group		Incident-beam direction					
		[100]		[110]		[*uuw*]	
215	*P*$\bar{4}$3*m*						
216	*F*$\bar{4}$3*m*						
217	*I*$\bar{4}$3*m*						
218	*P*$\bar{4}$3*n*	$h_okk, h_o\bar{k}k$ *n*	*A*	hhl_o *n*	*A*	hhl_o *n*	A_h
219	*F*$\bar{4}$3*c*	$h_ok_ok_o, h_o\bar{k}_ok_o$ *a*	*A*	$h_oh_ol_o$ *c*	*A*	$h_oh_ol_o$ *c*	A_h
220	*I*$\bar{4}$3*d*	$h_ekk, h_e\bar{k}k$: $2k + h_e = 4n + 2$ *d*	*A*	hhl_e: $2h + l_e = 4n + 2$ *d*	*A*	hhl_e: $2h + l_e = 4n + 2$ *d*	A_h

Table 2.5.3.12 (*cont.*)

Point group $m3m$. The symbol a in the column [100] is equivalent to the symbol c in the space groups of the first column.

Space group		Incident-beam direction							
		[100]		[110]		[uv0]		[uuw]	
221	$Pm\bar{3}m$ $P4/m\bar{3}2/m$								
222	$Pn\bar{3}n$ $P4/n\bar{3}2/n$	$h0l$: $h + l = 2n + 1$ n_{12} $hk0$: $h + k = 2n + 1$ n_{13} $h_{\rm o}kk, h_{\rm o}\bar{k}k$ n_2	A	$hk0$: $h + k = 2n + 1$ n_{13} $hhl_{\rm o}$ n_2	A	$hk0$: $h + k = 2n + 1$ n_1	$A_{\rm h}$	$hhl_{\rm o}$ n_2	$A_{\rm h}$
223	$Pm\bar{3}n$ $P4_2/m\bar{3}2/n$	$h_{\rm o}kk, h_{\rm o}\bar{k}k$ n	A	$hhl_{\rm o}$ n	A			$hhl_{\rm o}$ n	$A_{\rm h}$
224	$Pn\bar{3}m$ $P4_2/n\bar{3}2/m$	$h0l$: $h + l = 2n + 1$ n_2 $hk0$: $h + k = 2n + 1$ n_3	A	$hk0$: $h + k = 2n + 1$ n_3	A	$hk0$: $h + k = 2n + 1$ n	$A_{\rm h}$		
225	$Fm\bar{3}m$ $F4/m\bar{3}2/m$								
226	$Fm\bar{3}c$ $F4/m\bar{3}2/c$	$h_{\rm o}k_{\rm o}k_{\rm o}, h_{\rm o}\bar{k}_{\rm o}k_{\rm o}$ a	A	$h_{\rm o}h_{\rm o}l_{\rm o}$ c	A			$h_{\rm o}h_{\rm o}l_{\rm o}$ c	$A_{\rm h}$
227	$Fd\bar{3}m$ $F4_1/d\bar{3}2/m$	$h_{\rm e}0l_{\rm e}$: $h_{\rm e} + l_{\rm e} = 4n + 2$ d_2 $h_{\rm e}k_{\rm e}0$: $h_{\rm e} + k_{\rm e} = 4n + 2$ d_3	A	$h_{\rm e}k_{\rm e}0$: $h_{\rm e} + k_{\rm e} = 4n + 2$ d_3	A	$h_{\rm e}k_{\rm e}0$: $h_{\rm e} + k_{\rm e} = 4n + 2$ d	$A_{\rm h}$		
228	$Fd\bar{3}c$ $F4_1/d\bar{3}2/c$	$h_{\rm e}0l_{\rm e}$: $h_{\rm e} + l_{\rm e} = 4n + 2$ d_2 $h_{\rm e}k_{\rm e}0$: $h_{\rm e} + k_{\rm e} = 4n + 2$ d_3 $h_{\rm o}k_{\rm o}k_{\rm o}, h_{\rm o}\bar{k}_{\rm o}k_{\rm o}$ a	A	$h_{\rm o}h_{\rm o}l_{\rm o}$ c $h_{\rm e}k_{\rm e}0$: $h_{\rm e} + k_{\rm e} = 4n + 2$ d_3	A	$h_{\rm e}k_{\rm e}0$: $h_{\rm e} + k_{\rm e} = 4n + 2$ d	$A_{\rm h}$	$h_{\rm o}h_{\rm o}l_{\rm o}$ c	$A_{\rm h}$
229	$Im\bar{3}m$ $I4/m\bar{3}2/m$								
230	$Ia\bar{3}d$ $I4_1/a\bar{3}2/d$	$h_{\rm o}k_{\rm o}0$ a_3 $h_{\rm o}0l_{\rm o}$ c_2 $h_{\rm e}kk, h_{\rm e}\bar{k}k$: $2k + h_{\rm e} = 4n + 2$ d	A	$hhl_{\rm e}$: $2h + l_{\rm e} = 4n + 2$ d $h_{\rm o}k_{\rm o}0$ a_3	A	$h_{\rm o}k_{\rm o}0$ a	$A_{\rm h}$	$hhl_{\rm e}$: $2h + l_{\rm e} = 4n + 2$ d	$A_{\rm h}$

are no kinematically forbidden reflections. Thus, the lattice type is determined to be primitive P.

Possible space groups which satisfy point group $4/mmm$ and primitive lattice type P are those of Nos. 123–138 in Table 2.5.3.9. In Fig. 2.5.3.15(a), the dynamical extinction line A_2 is seen in the 100 disc and in the equivalent 010 disc. By consulting Table 2.5.3.9, four space groups $P4/mbm$, $P4/mnc$, $P4_2/mbc$ and $P4_2/mnm$ are selected. In Fig. 2.5.3.15(b), the dynamical extinction line A_2 is seen in the 010 disc but not in the $10\bar{1}$ disc. Two space groups $P4/mnc$ and $P4_2/mnm$ are selected from the four. To distinguish the two space groups, it is found from Table 2.5.3.9 that a CBED pattern taken with the [110] electron incidence should be examined. Fig. 2.5.3.15(e) shows a CBED pattern taken with the [110] incidence at 100 kV, where the 001 reflection is exactly excited. The $hh\bar{1}$ reflections are kinematically allowed for space group $P4_2/mnm$ but kinematically forbidden for $P4/mnc$. Since in the case of $P4/mnc$, no *Umweganregung* (multiple scattering) paths to the 001 reflection exist in the zeroth-order Laue zone, only the intensities of HOLZ lines, which are caused by *Umweganregung via* HOLZ reflections, are expected to appear in the 001 disc. If such *Umweganregung* is not practically excited, the 001 reflection must have no intensity. However, strong intensity produced by two-dimensional interaction is seen in the 001 disc of Fig. 2.5.3.15(e). This indicates that the reflection is an allowed reflection. Therefore, the space group of rutile is determined to be $P4_2/mnm$, which agrees with the space group already known.

Samarium selenide (Sm_3Se_4). Sm_3Se_4 has the Th_3P_4 structure type with space group $I\bar{4}3d$ at high temperatures. The lattice parameters are $a = b = c = 0.8885$ nm. It was expected that Sm_3Se_4 would transform to an ordered state of electrons with two valences of +2 and +3 around 150 K. The determination of the space group of the material was conducted at 100 K and room temperature. The space groups at both temperatures were determined by CBED to be the same. The following experiments were performed at 100 K.

Fig. 2.5.3.16(a) shows a CBED pattern taken with the [111] incidence at 80 kV, which clearly shows the first-order-Laue-zone reflections. The symmetry of the WP is seen to be $3m$ with the help of the enlarged insets. Possible diffraction groups are $3m$, $3m1_R$ and 6_Rmm_R from Table 2.5.3.3. Fig. 2.5.3.16(b), which is the central part of Fig. 2.5.3.16(a), shows projection symmetry $3m$, indicating that the projection diffraction group is $3m1_R$. Among the three groups $3m$, $3m1_R$ and 6_Rmm_R, diffraction groups for which the projection diffraction group is $3m1_R$ are $\bar{3}m$ and $3m1_R$. Possible point groups are found to be $3m$, $\bar{4}3m$ and $\bar{6}m2$ from Fig. 2.5.3.4. Fig. 2.5.3.16(c) shows a CBED pattern taken with the [100] incidence at 80 kV. The WP is seen to have symmetry $2mm$. Allowed diffraction groups are $2mm$, $2mm1_R$ and 4_Rmm_R. Fig. 2.5.3.16(d), which is the central part of Fig. 2.5.3.16(c), shows projection WP symmetry $4mm$, indicating that the projection diffraction group is $4mm1_R$. The diffraction group among the three groups $2mm$, $2mm1_R$ and 4_Rmm_R whose projection diffraction group is $4mm1_R$ is 4_Rmm_R. Possible point groups are found to be $\bar{4}3m$ and $\bar{4}2m$ from Fig. 2.5.3.4. Thus, the point group

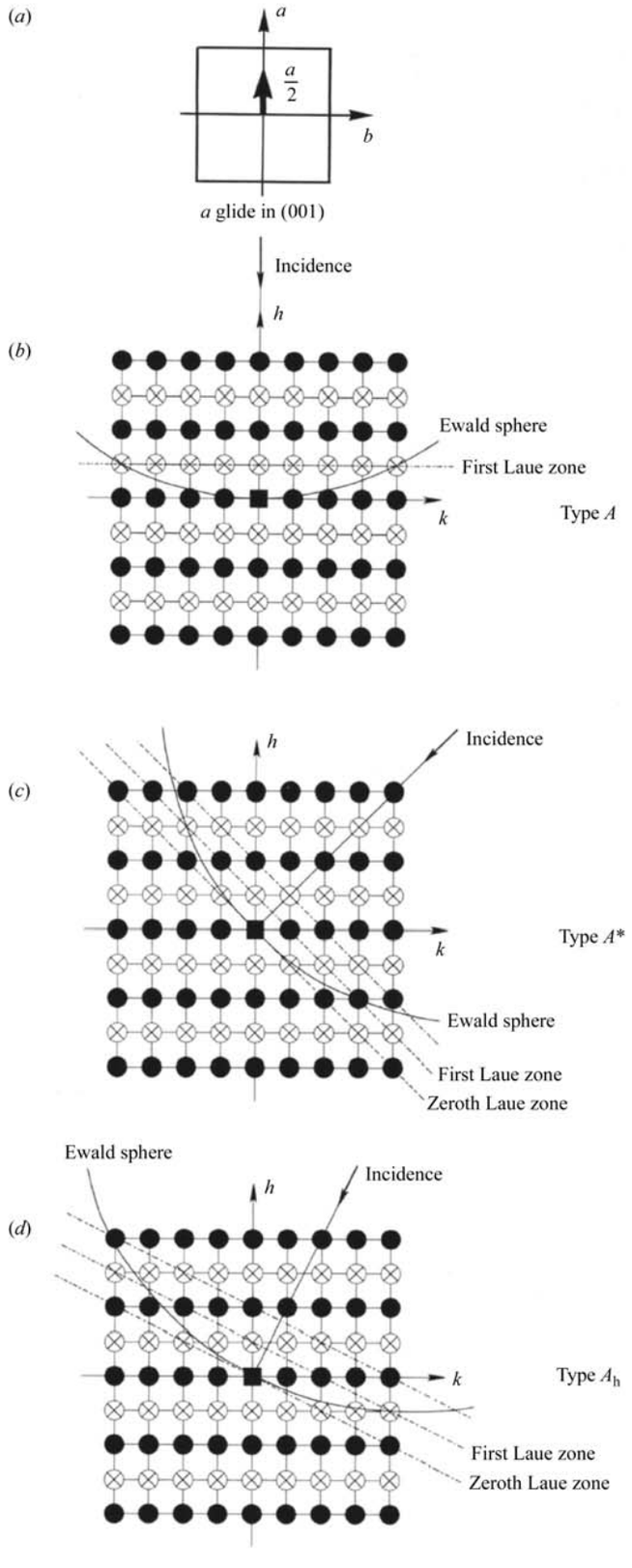


Fig. 2.5.3.13. Illustration of dynamical extinction lines appearing in HOLZ reflections due to glide planes. Black circles and circled crosses show kinematically allowed and kinematically forbidden reflections, respectively. (*a*) *a* glide in the (001) plane. (*b*) [100] incidence: dynamical extinction lines are formed in HOLZ reflections on both sides of the incident beam (type *A*). (*c*) [110] incidence: an extinction line is formed at a HOLZ reflection on one side of the incident beam because on the other side the Ewald sphere intersects an allowed HOLZ reflection (type *A**). (*d*) An incidence between [100] and [110]: an extinction line is formed at a HOLZ reflection on one side of the incident beam because on the other side the Ewald sphere does not intersect a HOLZ reflection (type A_h).

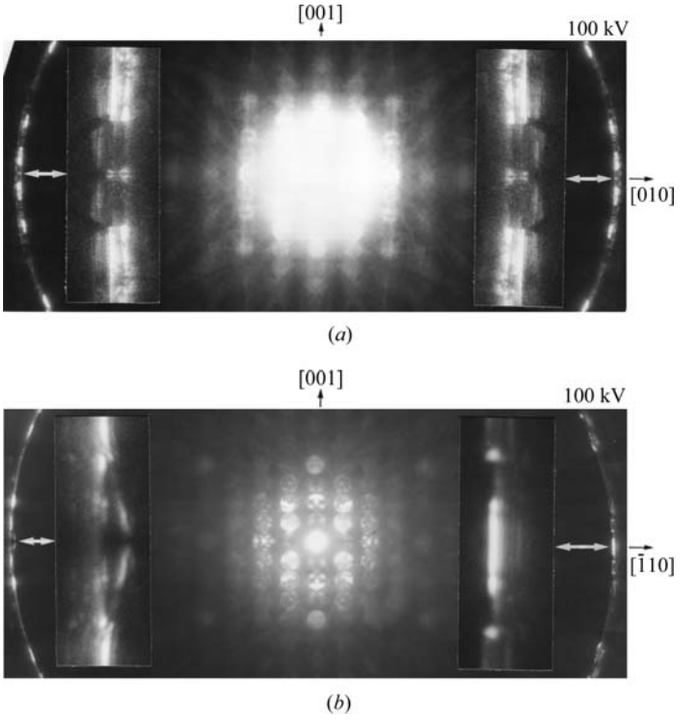


Fig. 2.5.3.14. HOLZ CBED pattern of FeS_2. (*a*) [100] incidence: type *A* dynamical extinction lines are seen clearly in the enlarged insets. (*b*) [110] incidence: a type *A** dynamical extinction line is seen clearly in the enlarged insets.

which satisfies the results obtained at the two crystal orientations is $\bar{4}3m$.

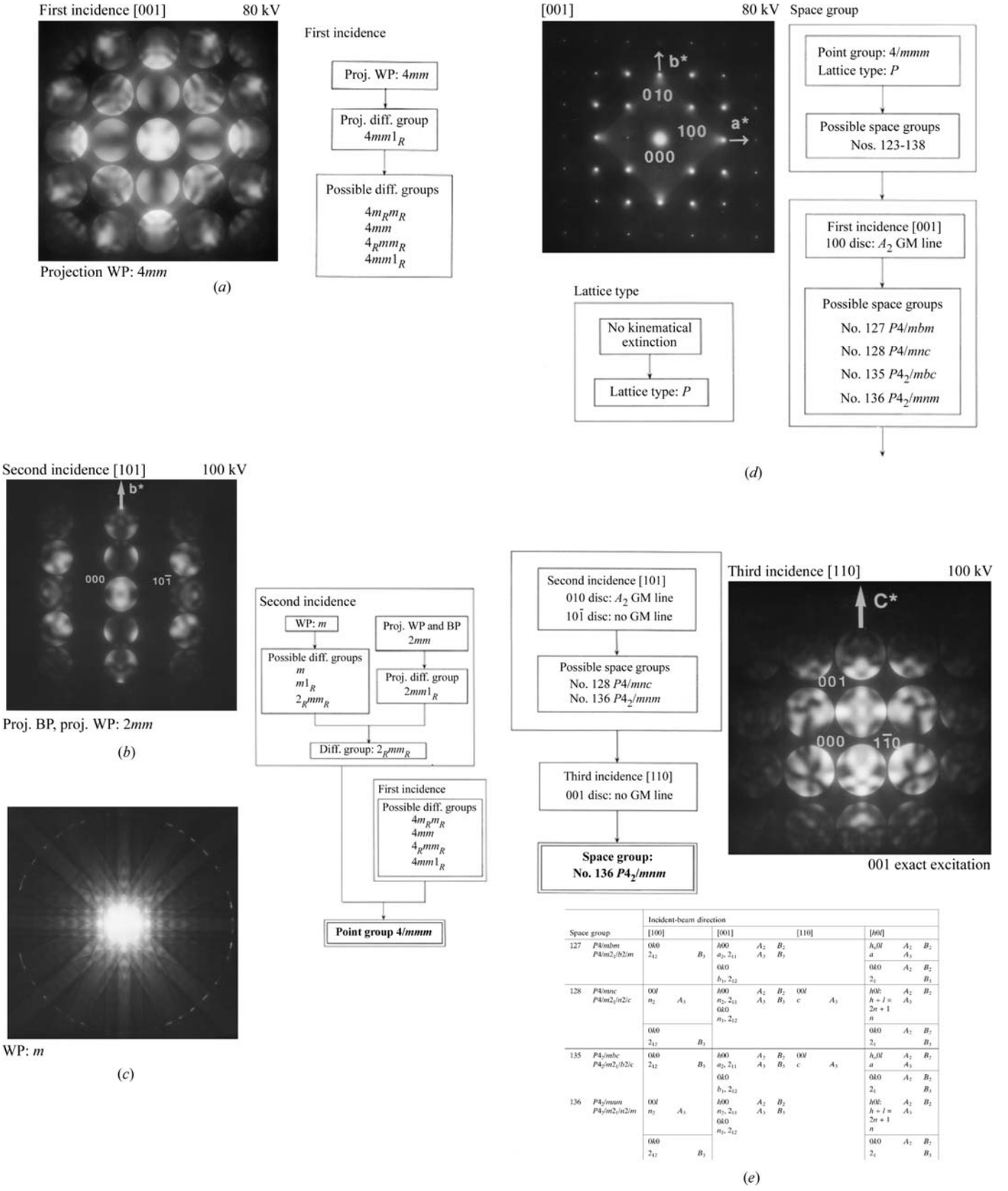
Fig. 2.5.3.16(*e*) shows an ordinary diffraction pattern taken with the [100] incidence at 80 kV. With the help of the lattice parameters and the camera length, the indices of the reflections are given as shown in the figure. The reflections $0kl$ ($k + l = 2n + 1$) are found to be kinematically forbidden. Thus, the lattice type is determined to be *I*.

The space groups having point group $\bar{4}3m$ and lattice type *I* are $I\bar{4}3m$ and $I\bar{4}3d$ from Table 2.5.3.9. Fig. 2.5.3.16(*d*) shows dynamical extinction lines A_2 in the 033 disc and equivalent discs (also broad lines A_2 in the 011 discs). Since the former space group does not give any dynamical extinction lines, the space group is determined to be $I\bar{4}3d$. For confirmation, a CBED pattern which contains the second-order-Laue-zone reflections was taken (Fig. 2.5.3.16*f*). Dynamical extinction lines *A* are seen in the 2,22,22 disc and the equivalent discs. This result also identifies the space group to be not $I\bar{4}3m$ but $I\bar{4}3d$ with the aid of Table 2.5.3.12.

2.5.3.4. Symmetry determination of incommensurate crystals

2.5.3.4.1. General remarks

Incommensurately modulated crystals do not have three-dimensional lattice periodicity. The crystals, however, recover lattice periodicity in a space higher than three dimensions. de Wolff (1974, 1977) showed that one-dimensional displacive and substitutionally modulated crystals can be described as a three-dimensional section of a (3 + 1)-dimensional periodic crystal. Janner & Janssen (1980*a*,*b*) developed a more general approach for describing a modulated crystal with *n* modulations as (3 + *n*)-dimensional periodic crystals ($n = 1, 2, \ldots$). Yamamoto (1982) derived a general structure-factor formula for *n*-dimensionally modulated crystals ($n = 1, 2, \ldots$), which holds for both displacive and substitutionally modulated crystals. Tables of the (3 + 1)-dimensional space groups for one-dimensional incommensurately modulated crystals were given by de Wolff *et al.* (1981), where the wavevector of the modulation was assumed to lie in the



Space group		Incident-beam direction [100]		[001]				[110]			[h0l]		
127	P4/mbm P4/m2$_1$/b2/m	0k0 2_{12}	B_3	h00 $a_2, 2_{11}$ 0k0 $b_1, 2_{12}$	A_2 A_3	B_2 B_3				h_n0l a 0k0 2_1	A_2 A_3 A_2	B_2 B_3 B_2 B_3
128	P4/mnc P4/m2$_1$/n2/c	00l n_2	A_3	h00 $n_2, 2_{11}$ 0k0 $n_1, 2_{12}$	A_2 A_3	B_2 B_3	00l c	A_3		h0l: $h + l =$ $2n + 1$ n	A_2 A_3	B_2
		0k0 2_{12}	B_3							0k0 2_1	A_2	B_2 B_3
135	P4$_2$/mbc P4$_2$/m2$_1$/b2/c	0k0 2_{12}	B_3	h00 $a_2, 2_{11}$ 0k0 $b_1, 2_{12}$	A_2 A_3	B_2 B_3	00l c	A_3		h_n0l a 0k0 2_1	A_3 A_3 A_2	B_2 B_3 B_2 B_3
136	P4$_2$/mnm P4$_2$/m2$_1$/n2/m	00l n_2	A_3	h00 $n_2, 2_{11}$ 0k0 $n_1, 2_{12}$	A_2 A_3	B_2 B_3				h0l: $h + l =$ $2n + 1$ n	A_2 A_3	B_2
		0k0 2_{12}	B_3							0k0 2_1	A_2	B_2 B_3

Fig. 2.5.3.15. CBED patterns of rutile. The procedures for identifying the symmetry are also shown. (*a*) [001] incidence at 80 kV: the projection (proj.) WP shows symmetry 4*mm*. (*b*, *c*) [101] incidence at 100 kV: the projection BP and projection WP show symmetry 2*mm* and the WP shows symmetry *m* (the point group is 4/*mmm*). (*d*) Spot diffraction pattern showing no extinction caused by the lattice type (lattice type *P*). (*e*) Near-[110] incidence at 100 kV to excite exactly the 001 reflection: no extinction lines in the 001 disc (space group $P4_2/mnm$).

c direction. Later, some corrections to the tables were made by Yamamoto *et al.* (1985). The analysis of incommensurately modulated crystals using (3 + 1)-dimensional space groups has become familiar in the field of X-ray structure analysis.

Fung *et al.* (1980) applied the CBED method to the study of incommensurately modulated transition-metal dichalcogenides. Steeds *et al.* (1985) applied the LACBED method (Tanaka *et al.*, 1980) to the study of incommensurately modulated crystals of $NiGe_{1-x}P_x$. Tanaka *et al.* (1988, pp. 74–81) examined the symmetries of the incommensurate and fundamental reflections appearing in the CBED patterns obtained from the incommensurately modulated crystals of $Sr_2Nb_2O_7$ and Mo_8O_{23}. Terauchi &

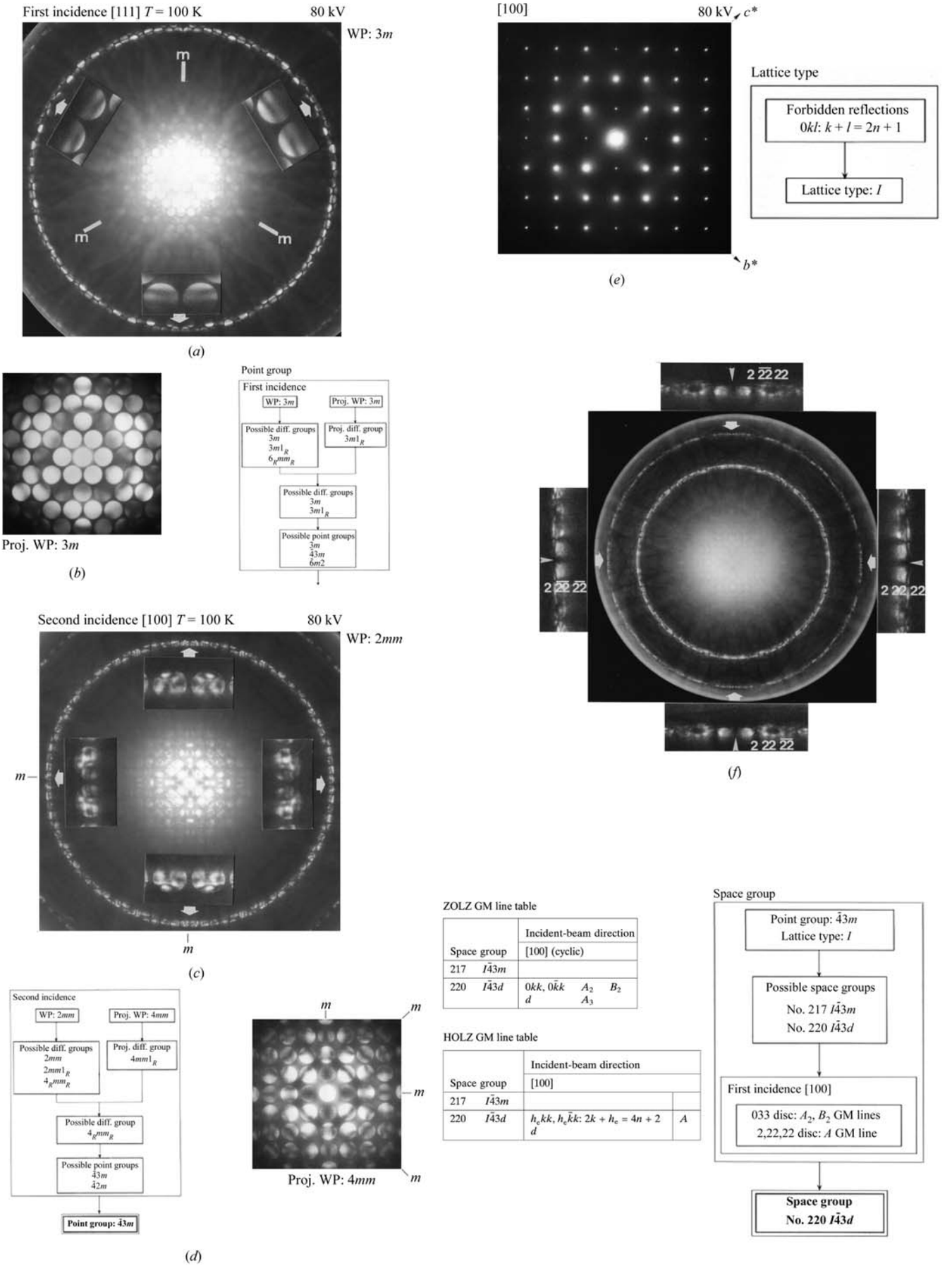


Fig. 2.5.3.16. CBED patterns of Sm_3Se_4. The procedures for identifying the symmetry are also shown. (*a*, *b*) [111] incidence at 80 kV: the WP symmetry is 3*m* (*a*) and the projection (proj.) WP symmetry is 3*m* (*b*). (*c*, *d*) [100] incidence at 80 kV: the WP symmetry is 2*mm* (*c*) and the projection WP symmetry is 4*mm* (*d*). Dynamical extinction lines A_2 and A_3 are seen (*d*). The point group is determined to be $\bar{4}3m$. (*e*) Spot diffraction pattern taken with the [100] incidence at 80 kV shows the absence of 0*kl* reflections. The lattice type is determined to be *I*. (*f*) [100] incidence at 100 kV: dynamical extinction lines *A* in HOLZ reflections confirm the existence of a glide plane. The space group is determined to be $I\bar{4}3d$.

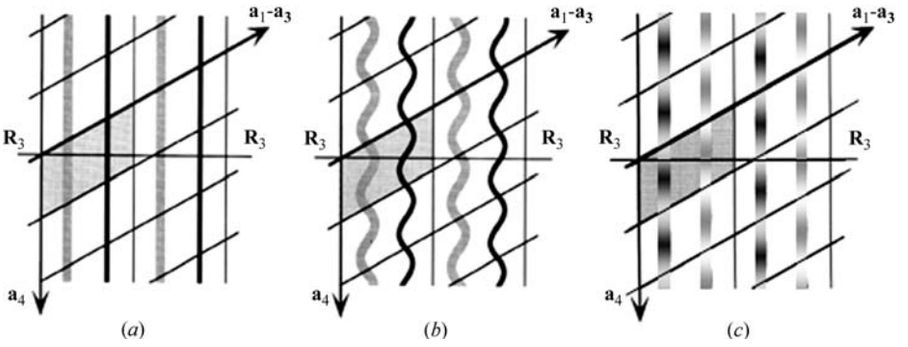
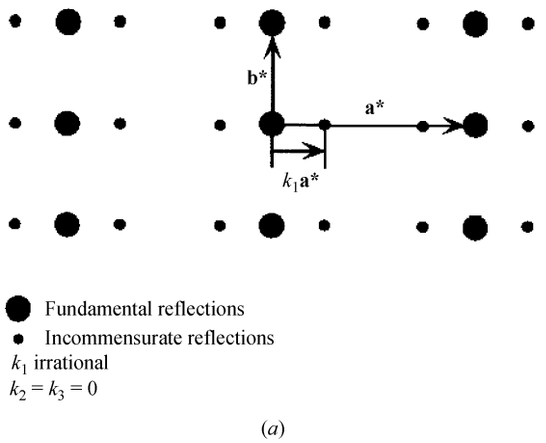


Fig. 2.5.3.17. The (3 + 1)-dimensional description of one-dimensionally modulated crystals. Atoms are shown as strings along the fourth direction $\mathbf{a}_4$. (*a*) No modulation, shown as straight strings. (*b*) Displacive modulation, shown as wavy strings. (*c*) Amplitude modulation with varying-density strings.

Tanaka (1993) clarified theoretically the interrelation between the symmetries of CBED patterns and the (3 + 1)-dimensional point-group symbols for incommensurately modulated crystals and verified experimentally the theoretical results for $Sr_2Nb_2O_7$ and Mo_8O_{23}. Terauchi *et al.* (1994) investigated dynamical extinction for the (3 + 1)-dimensional space groups. They clarified that approximate dynamical extinction lines appear in CBED discs of the reflections caused by incommensurate modulations when the amplitudes of the incommensurate modulation waves are small. They tabulated the dynamical extinction lines appearing in the CBED discs for all the (3 + 1)-dimensional space groups of the incommensurately modulated crystals. The tables were stored in the British Library Document Supply Centre as Supplementary Publication No. SUP 71810 (65 pp.). They showed an example of the dynamical extinction lines obtained from $Sr_2Nb_2O_7$. The point- and space-group determinations of the (3 + 1)-dimensional crystals are described compactly in the book by Tanaka *et al.* (1994, pp. 156–205).



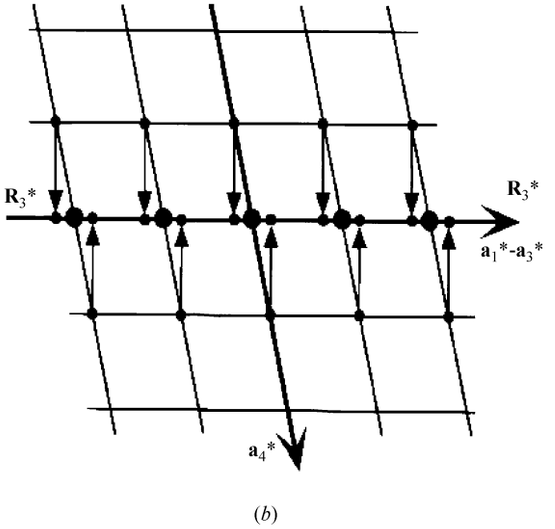


Fig. 2.5.3.18. (*a*) Schematic diffraction pattern from a modulated crystal. As an example, the wave number vector of modulation is assumed to be $k_1\mathbf{a}^*$, k_1 being an irrational number. Large and small spots denote fundamental and incommensurate reflections, respectively. (*b*) Incommensurate reflections are obtained by a projection of the Fourier transform of a (3 + 1)-dimensional periodic structure.

Fig. 2.5.3.17 illustrates (3 + 1)-dimensional descriptions of a crystal structure without modulation (*a*), a one-dimensional displacive modulated structure (*b*) and a one-dimensional substitutionally modulated structure (*c*). The arrows labelled $\mathbf{a}_1$–$\mathbf{a}_3$ (**a**, **b** and **c**) and $\mathbf{a}_4$ indicate the (3 + 1)-dimensional crystal axes. The horizontal line labelled $\mathbf{R}_3$ represents the three-dimensional space (external space). In the (3 + 1)-dimensional description, an atom is not located at a point as in the three-dimensional space, but extends as a string along the fourth direction $\mathbf{a}_4$ perpendicular to the three-dimensional space $\mathbf{R}_3$. The shaded parallelogram is a unit cell in the (3 + 1)-dimensional space. The unit cell contains two atom strings in this case. In the case of no modulations, the atoms are shown as straight strings, as shown in Fig. 2.5.3.17(*a*). For a displacive modulation, atoms are expressed by wavy strings periodic along the fourth direction $\mathbf{a}_4$ as shown in Fig. 2.5.3.17(*b*). The width of the atom strings indicates the spread of the atoms in $\mathbf{R}_3$. The atom positions of the modulated structure in $\mathbf{R}_3$ are given as a three-dimensional ($\mathbf{R}_3$) section of the atom strings in the (3 + 1)-dimensional space. A substitutional modulation, which is described by a modulation of the atom form factor, is expressed by atom strings with a density modulation along the direction $\mathbf{a}_4$ as shown in Fig. 2.5.3.17(*c*).

The diffraction vector **G** is written as

$$\mathbf{G} = h_1\mathbf{a}^* + h_2\mathbf{b}^* + h_3\mathbf{c}^* + h_4\mathbf{k},$$

where a set of $h_1h_2h_3h_4$ is a (3 + 1)-dimensional reflection index, and **a***, **b*** and **c*** are the reciprocal-lattice vectors of the real-lattice vectors **a**, **b** and **c** of the average structure. The modulation vector **k** is written as

$$\mathbf{k} = k_1\mathbf{a}^* + k_2\mathbf{b}^* + k_3\mathbf{c}^*,$$

where one coefficient k_i (i = 1–3) is an irrational number and the others are rational. Fig. 2.5.3.18(*a*) shows a diffraction pattern of a crystal with an incommensurate modulation wavevector $k_1\mathbf{a}^*$ (k_2 and k_3 = 0). Large and small black spots show the fundamental reflections and incommensurate reflections, respectively, only the first-order incommensurate reflections being shown. It

should be noted that the diffraction pattern of a modulated crystal is obtained by a projection of the Fourier transform of the (3 + 1)-dimensional periodic structure. Fig. 2.5.3.18(*b*) is assumed to be the Fourier transform of Fig. 2.5.3.17(*b*). Incommensurate reflections are obtained by a projection of the reciprocal-lattice points onto $\mathbf{R}_3^*$.

The displacive modulation is expressed by the atom displacement u_i^μ with x_4. The structure factor $F(h_1h_2h_3h_4)$ for the (3 + 1)-dimensional crystal with a displacive modulation is given by de Wolff (1974, 1977) as follows:

$$F(h_1h_2h_3h_4) = \sum_{\mu=1}^{N} f_\mu \exp[2\pi i(h_1\bar{x}_1^\mu + h_2\bar{x}_2^\mu + h_3\bar{x}_3^\mu)] \times \int_0^1 \exp\left\{2\pi i\left[\sum_{i=1}^{3}(h_i + h_4k_i)u_i^\mu + h_4\bar{x}_4^\mu\right]\right\} \mathrm{d}\bar{x}_4^\mu, \tag{2.5.3.6}$$

where

$$\bar{x}_4^\mu = (\bar{x}_1^\mu + n_1)k_1 + (\bar{x}_2^\mu + n_2)k_2 + (\bar{x}_3^\mu + n_3)k_3.$$

f_μ and $\bar{x}_i^\mu (i = 1\text{–}3)$ are, respectively, the atom form factor and the ith component of the position of the μth atom in the unit cell of the average structure. The symbol u_i^μ is the ith component of the displacement of the μth atom. Since the atom in the (3 + 1)-dimensional space is continuous along $\mathbf{a}_4$ and discrete along $\mathbf{R}_3$, the structure factor is expressed by summation in $\mathbf{R}_3$ and integration along $\mathbf{a}_4$ as seen in equation (2.5.3.6). The integration implies that the sum for the atoms with displacements is taken over the infinite number of unit cells of the average structure. That is, equation (2.5.3.6) is the structure factor for a unit cell with the lattice parameter of an infinite length in $\mathbf{R}_3$ along the direction of the modulation wavevector $\mathbf{k}$.

CBED patterns are obtained from a finite area of a specimen crystal. For the symmetry analysis of CBED patterns obtained from modulated structures, the effect of the finite size was considered by Terauchi & Tanaka (1993). The integration over a unit-cell length along $\mathbf{a}_4$ in equation (2.5.3.6) is rewritten in the following way with the summation over a finite number of three-dimensional sections of the atom strings:

$$F'(h_1h_2h_3h_4) = \sum_{\mu=1}^{N} f_\mu \exp[2\pi i(h_1\bar{x}_1^\mu + h_2\bar{x}_2^\mu + h_3\bar{x}_3^\mu)] \times \sum_{n_1}\sum_{n_2}\sum_{n_3} \exp\left\{2\pi i\left[\sum_{i=1}^{3}(h_i + h_4k_i)u_i^\mu + h_4\bar{x}_4^\mu\right]\right\}, \tag{2.5.3.7}$$

where $N_1 < n_1 \le N_1'$, $N_2 < n_2 \le N_2'$ and $N_3 < n_3 \le N_3'$, $N' = (N_1' - N_1)(N_2' - N_2)(N_3' - N_3)$ being the number of unit cells of the average structure included in a specimen volume from which CBED patterns are taken.

The substitutional modulation arises from a periodic variation of the site-occupation probability of the atoms. This modulation is expressed by a modulation of the atom form factor f_μ with x_4. The structure factor $F'(h_1h_2h_3h_4)$ for a finite-size crystal is written as

$$F'(h_1h_2h_3h_4) = \sum_{\mu=1}^{N} \exp[2\pi i\,(h_1x_1^\mu + h_2x_2^\mu + h_3x_3^\mu)] \times \sum_{n_1}\sum_{n_2}\sum_{n_3} f_\mu \exp(2\pi i h_4 x_4^\mu), \tag{2.5.3.8}$$

where $x_4^\mu = \sum_i (x_i^\mu + n_i)k_i$.

2.5.3.4.2. Point-group determination

The symmetries of the CBED patterns can be determined by examination of the symmetries of the structure factor $F'(h_1h_2h_3h_4)$ in equations (2.5.3.7) or (2.5.3.8). We consider a displacive modulated structure, which has a modulation wavevector $\mathbf{k} = k_3\mathbf{c}^*$ and belongs to (3 + 1)-dimensional space group $P^{P2/m}_{1\ \bar{1}}$. This space-group symbol implies the following.

(1) The modulation wavevector $\mathbf{k}$ exists inside the first Brillouin zone of the average structure (P).

(2) The average structure belongs to space group $P2/m$, the twofold rotation axis being parallel to the c axis.

(3) The symmetry subsymbol 1, which is written beneath symmetry symbol 2, indicates that the modulation wavevector $\mathbf{k}$ is transformed into itself by symmetry operation 2 of the average structure. The symmetry subsymbol beneath symmetry symbol m indicates that the wavevector $\mathbf{k}$ is transformed into $-\mathbf{k}$ by symmetry operation m. The modulated structure has a twofold rotation axis, which is common to the average structure, but does not have mirror symmetry, which is possessed by the average structure.

For the twofold rotation axis (symbol 2) of this space group, the structure factor $F'(h_1h_2h_3h_4)$ is written as

$$\begin{aligned} F'(h_1h_2h_3h_4) &= \sum_{\mu=1}^{N} f_\mu \exp[2\pi i(h_1\bar{x}_1^\mu + h_2\bar{x}_2^\mu + h_3\bar{x}_3^\mu)] \\ &\quad \times \sum_{n_3} \exp\{2\pi i[h_1u_1^\mu + h_2u_2^\mu + (h_3 + h_4k_3)u_3^\mu + h_4\bar{x}_4^\mu]\} \\ &\quad + \sum_{\mu=1}^{N} f_\mu \exp[2\pi i(-h_1\bar{x}_1^\mu - h_2\bar{x}_2^\mu + h_3\bar{x}_3^\mu)] \\ &\quad \times \sum_{n_3} \exp\{2\pi i[-h_1u_1^\mu - h_2u_2^\mu + (h_3 + h_4k_3)u_3^\mu + h_4\bar{x}_4^\mu]\}, \end{aligned} \tag{2.5.3.9}$$

where $x_4^\mu = (x_3^\mu + n_3)k_3$. It is found from equation (2.5.3.9) that two structure factors $F'(h_1h_2h_3h_4)$ and $F'(\bar{h}_1\bar{h}_2h_3h_4)$ are the same when reflections $h_1h_2h_3h_4$ and $\bar{h}_1\bar{h}_2h_3h_4$ are equivalent with respect to the twofold rotation axis of the average structure. Thus, not only fundamental reflections ($h_4 = 0$) from the average structure but also the satellite reflections ($h_4 \neq 0$) from the incommensurate structure show twofold rotation symmetry about the c^* axis.

For the mirror plane (symbol m), the structure factor is written in a similar manner to the case of the twofold rotation axis. It is found that $F'(h_1h_2h_3h_4)$ is not equal to $F'(h_1h_2\bar{h}_3\bar{h}_4)$ for the incommensurate reflections $h_4 \neq 0$. Hence, the incommensurate reflections do not show mirror symmetry with respect to the mirror plane of the average structure. For the fundamental reflections ($h_4 = 0$), $F'(h_1h_2h_3h_4)$ is equal to $F'(h_1h_2\bar{h}_3\bar{h}_4)$, indicating the existence of mirror symmetry. It should be noted that the mirror symmetry can be destroyed by the dynamical diffraction effect between the fundamental and incommensurate reflections. In most modulated structures, however, the amplitude of the modulation wave u_i^μ is not so large as to destroy the symmetry of the fundamental reflections. Thus, the fundamental

Table 2.5.3.13. *Wavevectors, point- and space-group symbols and CBED symmetries of one-dimensionally modulated crystals*

Wavevector transformation	Point-group symbol	Symmetry of incommensurate reflection	Space-group symbol	Dynamical extinction lines
$\mathbf{k} \rightarrow \mathbf{k}$	1	Same symmetry as average structure	1, s (1/2), t ($\pm$1/3), q ($\pm$1/4), h ($\pm$1/6)	Yes for s, q and h
$\mathbf{k} \rightarrow -\mathbf{k}$	$\bar{1}$	No symmetry	$\bar{1}$	No

reflections ought to show the symmetry of the average structure, while the incommensurate reflections lose the symmetry.

The problem of the finite size of the illuminated area is discussed using equations (2.5.3.7) and (2.5.3.8) in a paper by Terauchi & Tanaka (1993) and in the book by Tanaka *et al.* (1994, pp. 156–205). The results are as follows: Even if the size and position of an illuminated specimen area are changed, the intensity distribution in a CBED pattern changes but the symmetry of the pattern does not. To obtain the symmetries of incommensurate crystals, it is not necessary to take CBED patterns from an area whose diameter is larger than the period of the modulated structure. The symmetries of the modulated structure can appear when more than one unit cell of the average structure is illuminated for displacive modulations. For substitutional modulations, a specimen volume that produces the average atom form factor is needed, namely a volume of about 1 nm diameter area and 50 nm thick.

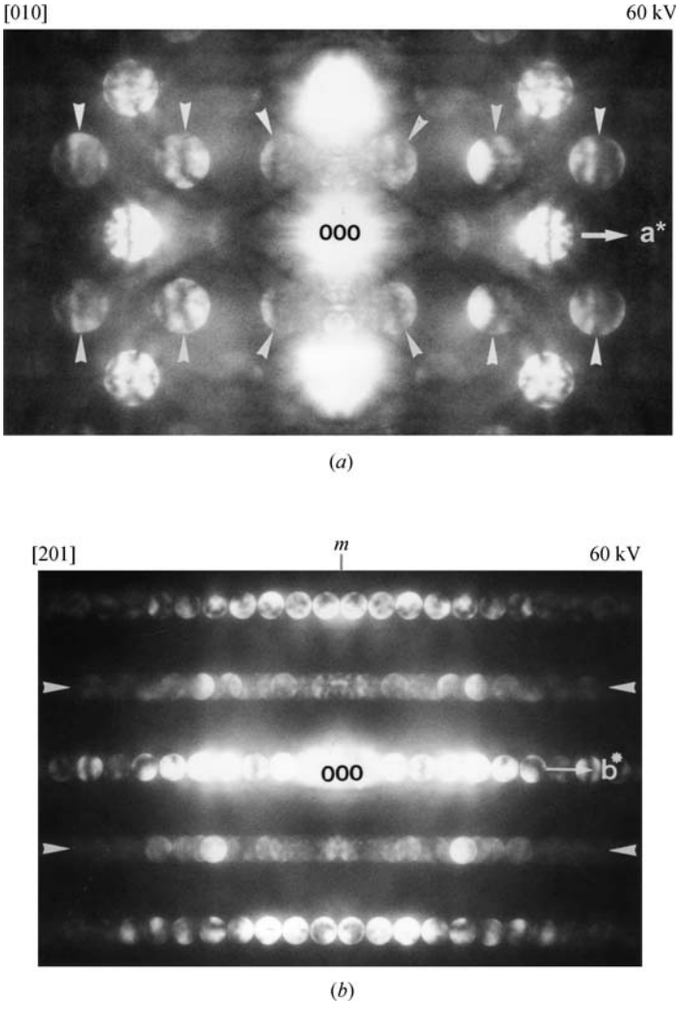


Fig. 2.5.3.19. CBED patterns of the incommensurate phase of $Sr_2Nb_2O_7$ taken at 60 kV. (*a*) [010] incidence: fundamental reflections show a mirror symmetry perpendicular to the a^* axis but incommensurate reflections do not [symmetry ($^m_1$)]. (*b*) [201] incidence: incommensurate reflections show mirror symmetry perpendicular to the b^* axis [symmetry ($^m_1$)]. The wave number vector of the modulation is $\mathbf{k} = (\frac{1}{2} - \delta)\mathbf{a}^*$.

Table 2.5.3.13 shows the point-group symmetries (third column) of the incommensurate reflections for the two point-group subsymbols. For symmetry subsymbol 1, both the fundamental and incommensurate reflections show the symmetries of the average structure. For symmetry subsymbol $\bar{1}$, the fundamental reflections show the symmetries of the average structure but the incommensurate reflections do not have any symmetry. These facts imply that the symmetries of the incommensurate reflections are determined by the point group of the average structure and the modulation wavevector **k**. In other words, observation of the symmetries of the incommensurate reflections is not necessary for the determination of the point groups, although it can ascertain the point groups of the modulated crystals.

An example of point-group determination is shown for the incommensurate phase of $Sr_2Nb_2O_7$. Many materials of the $A_2B_2O_7$ family undergo phase transformations from space group $Cmcm$ to $Cmc2_1$ and further to $P2_1$ with decreasing temperature. An incommensurate phase appears between the $Cmc2_1$ phase and the $P2_1$ phase. $Sr_2Nb_2O_7$ transforms at 488 K from the $Cmc2_1$ phase into the incommensurate phase with a modulation wavevector $\mathbf{k} = (\frac{1}{2} - \delta)\mathbf{a}^*$ (δ = 0.009–0.023) but does not transform into the $P2_1$ phase. The space group of $Sr_2Nb_2O_7$ was reported as $P^{C\,mc2_1}_{\bar{1}s\bar{1}}$ (Yamamoto, 1988). (Since the space-group notation $Cmc2_1$ is broadly accepted, the direction of the modulation is taken as the a axis.) The point group of the phase is $^{mm2}_{\bar{1}\,\bar{1}\,1}$. The modulation wavevector **k** is transformed to $-\mathbf{k}$ by the mirror symmetry operation perpendicular to the a axis ($^m_1$) and by the twofold rotation symmetry operation about the c axis ($^2_1$). The wavevector is transformed into itself by the mirror symmetry operation perpendicular to the b axis ($^m_1$).

Fig. 2.5.3.19(*a*) shows a CBED pattern of the incommensurate phase of $Sr_2Nb_2O_7$ taken with the [010] incidence at an accelerating voltage of 60 kV. The reflections indicated by arrowheads are the incommensurate reflections. Other reflections are the fundamental reflections. Since the pattern is produced by the interaction of the reflections in the zeroth-order Laue zone, symmetry operations ($^m_1$) and ($^2_1$) act the same. These symmetries are confirmed by the fact that the fundamental reflections show mirror symmetry perpendicular to the a^* axis (twofold rotation symmetry about the c^* axis) but the incommensurate reflections do not. Fig. 2.5.3.19(*b*) shows a CBED pattern of the incommensurate phase of $Sr_2Nb_2O_7$ taken with the [201] incidence at 60 kV. The reflections in the two rows indicated by arrowheads are the incommensurate reflections and the others are the fundamental reflections. Symmetry symbol ($^m_1$) implies that both the fundamental and incommensurate reflections display mirror symmetry perpendicular to the b^* axis. Fig. 2.5.3.19(*b*) exactly exhibits the symmetry.

2.5.3.4.3. Space-group determination

Table 2.5.3.13 shows the space-group symbols (fourth column) of the modulated crystals. When a glide (screw) component τ_4 between the modulation waves of two atom rows is 0, 1/2, $\pm$1/3, $\pm$1/4 or $\pm$1/6, symbol 1, s, t, q or h is given, respectively (de Wolff *et al.*, 1981). Such glide components are allowed for point-group symmetry 1 but are not for point-group symmetry $\bar{1}$. Dynamical

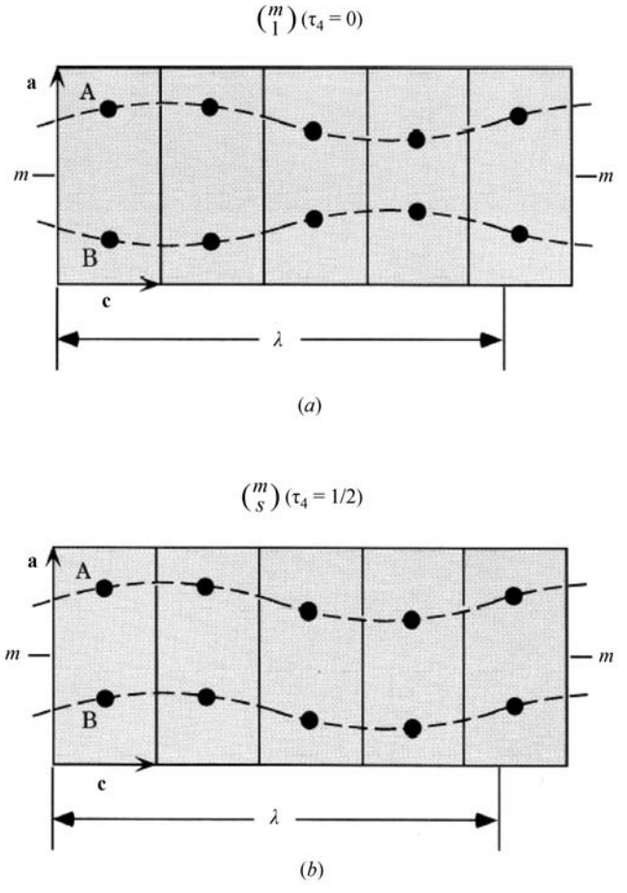


Fig. 2.5.3.20. (*a*) Mirror symmetry of modulation waves $\binom{m}{1}$ $\tau_4 = 0$. (*b*) Glide symmetry of modulation waves $\binom{m}{s}$ $\tau_4 = \frac{1}{2}$. The wave number vector of modulation is $k_3\mathbf{c}^*$.

extinction occurs for glide components *s*, *q* and *h* but does not for glide component *t*. When the average structure does not have a glide component, dynamical extinction due to a glide component τ_4 appears in odd-order incommensurate reflections. When the average structure has a glide component, dynamical extinction due to a glide component τ_4 appears in incommensurate reflections with $h_i + h_4 = 2n + 1$, where h_i and h_4 are the reflection indices for the average structure and incommensurate structure, respectively. Details are given in the paper by Terauchi *et al.* (1994).

Fig. 2.5.3.20(*a*) illustrates mirror symmetry $\binom{m}{1}$ between atom rows A and B, which is perpendicular to the *b* axis with no glide component ($\tau_4 = 0$). Here, the wave number vector of the modulation is assumed to be $\mathbf{k} = k_3\mathbf{c}^*$ following the treatment of de Wolff *et al.* (1981). Fig. 2.5.3.20(*b*) illustrates glide symmetry $\binom{m}{s}$ with a glide component $\tau_4 = \frac{1}{2}$. The structure factor $F(h_1h_2h_3h_4)$ is written for the glide plane $\binom{m}{s}$ of an infinite incommensurate crystal as

$$\begin{aligned}
&F(h_1h_2h_3h_4)\\
&= \sum_{\mu=1}^{N} f_\mu \exp[2\pi i(h_1\bar{x}_1^\mu + h_2\bar{x}_2^\mu + h_3\bar{x}_3^\mu)]\\
&\quad \times \int_0^1 \exp\{2\pi i[h_1u_1^\mu + h_2u_2^\mu + (h_3 + h_4k_3)u_3^\mu + h_4\bar{x}_4^\mu]\}\,\mathrm{d}\bar{x}_4^\mu\\
&\quad + \exp(h_4\pi i)\sum_{\mu=1}^{N} f_\mu \exp[2\pi i(h_1\bar{x}_1^\mu - h_2\bar{x}_2^\mu + h_3\bar{x}_3^\mu)]\\
&\quad \times \int_0^1 \exp\{2\pi i[h_1u_1^\mu - h_2u_2^\mu + (h_3 + h_4k_3)u_3^\mu + h_4\bar{x}_4^\mu]\}\,\mathrm{d}\bar{x}_4^\mu.
\end{aligned} \tag{2.5.3.10}$$

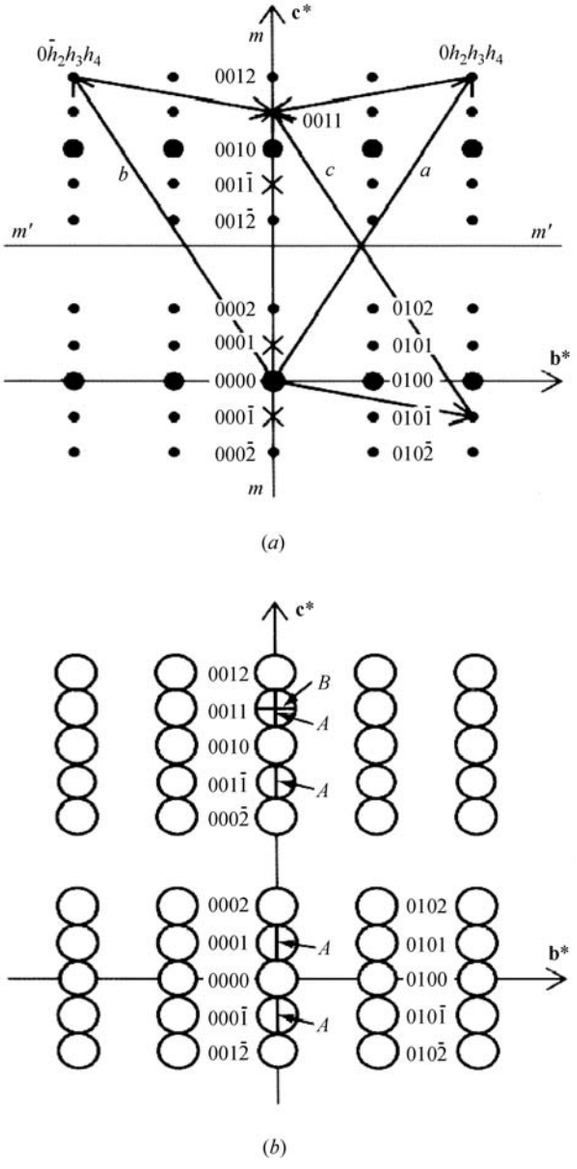


Fig. 2.5.3.21. (*a*) *Umweganregung* paths *a*, *b* and *c* to the 0011 forbidden reflection. (*b*) Expected dynamical extinction lines are shown, the 0011 reflection being excited. The wave number vector of modulation is $k_3\mathbf{c}^*$.

Thus, the following phase relations are obtained between the two structure factors:

$$\begin{aligned}
F(h_1h_2h_3h_4) &= F(h_1\bar{h}_2h_3h_4) \quad \text{for } h_4 \text{ even},\\
F(h_1h_2h_3h_4) &= -F(h_1\bar{h}_2h_3h_4) \quad \text{for } h_4 \text{ odd}.
\end{aligned} \tag{2.5.3.11}$$

These relations are analogous to the phase relations between the two structure factors for an ordinary three-dimensional crystal with a glide plane. The relations imply that dynamical extinction occurs for the glide planes and screw axes of the (3 + 1)-dimensional crystal with an infinite dimension along the direction of the incommensurate modulation wavevector **k**. Terauchi *et al.* (1994) showed that approximate dynamical extinction occurs for an incommensurate crystal of finite dimension.

Fig. 2.5.3.21(*a*) and (*b*) illustrate a spot diffraction pattern and a CBED pattern, respectively, expected from a modulated crystal with a (3 + 1)-dimensional space group $P^{P2mm}_{\bar{1}s\bar{1}}$ ($\mathbf{k} = k_3\mathbf{c}^*$) at the [100] incidence. The large and small spots in Fig. 2.5.3.21(*a*) designate the fundamental ($h_4 = 0$) and incommensurate reflections ($h_4 \neq 0$), respectively. The $00h_3h_4$ (h_4 = odd) reflections

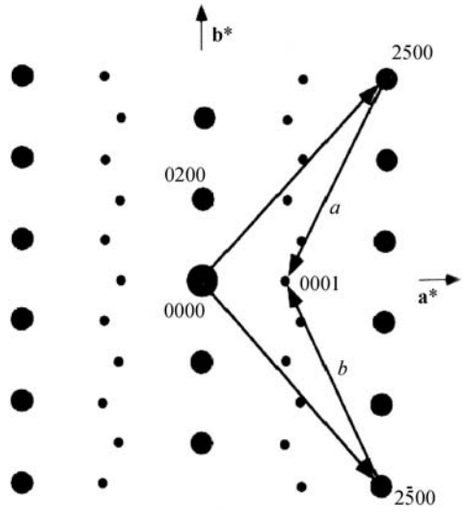


Fig. 2.5.3.22. Schematic diffraction pattern at the [001] incidence of $Sr_2Nb_2O_7$. *Umweganregung* paths *a* and *b via* fundamental reflections to the 0001 incommensurate reflection. Large and small spots denote fundamental and incommensurate reflections, respectively. The wave number vector of modulation is $\mathbf{k} = (\frac{1}{2} - \delta)\mathbf{a}^*$.

shown by crosses are kinematically forbidden by the glide plane (^m_s) perpendicular to the *b* axis. *Umweganregung* paths *a*, *b* and *c* in the ZOLZ to a kinematically forbidden reflection are drawn. The two paths *a* and *b* are geometrically equivalent with respect to the line *m*–*m* perpendicular to the *b* axis. Since every *Umweganregung* path to a kinematically forbidden reflection contains an odd number of $F(0h_{2,i}h_{3,i}h_{4,i})$ with odd $h_{4,i}$, the following equation is obtained.

$$F(0h_{2,1}h_{3,1}h_{4,1})F(0h_{2,2}h_{3,2}h_{4,2})\ldots F(0h_{2,n}h_{3,n}h_{4,n}) \quad \text{for path } a$$
$$= -F(0\bar{h}_{2,1}h_{3,1}h_{4,1})F(0\bar{h}_{2,2}h_{3,2}h_{4,2})\ldots F(0\bar{h}_{2,n}h_{3,n}h_{4,n})$$
$$\text{for path } b, \qquad (2.5.3.12)$$

where $\sum_{i=1}^{n} h_{2,i} = 0$, $\sum_{i=1}^{n} h_{3,i} = h_3$ and $\sum_{i=1}^{n} h_{4,i} = h_4$ (h_4 = odd).

When reflection $00h_3h_4$ (h_4 = odd) is exactly excited, the two paths *a* and *c* are symmetric with respect to the bisector m'–m' of the diffraction vector of the reflection and have the same excitation error. The waves passing through these paths have the same amplitude but different signs. Thus the following relation is obtained.

$$F(0h_{2,1}h_{3,1}h_{4,1})F(0h_{2,2}h_{3,2}h_{4,2})\ldots F(0h_{2,n}h_{3,n}h_{4,n}) \quad \text{for path } a$$
$$= -F(0\bar{h}_{2,n}h_{3,n}h_{4,n})F(0\bar{h}_{2,n-1}h_{3,n-1}h_{4,n-1})\ldots F(0\bar{h}_{2,1}h_{3,1}h_{4,1})$$
$$\text{for path } c, \qquad (2.5.3.13)$$

where $\sum_{i=1}^{n} h_{2,i} = 0$, $\sum_{i=1}^{n} h_{3,i} = h_3$ and $\sum_{i=1}^{n} h_{4,i} = h_4$ (h_4 = odd).

Therefore, dynamical extinction occurs in kinematically forbidden reflections of incommensurate crystals. Fig. 2.5.3.21(*b*) schematically shows the extinction lines in odd-order incommensurate reflections, where the 0011 reflection is exactly excited.

We consider the dynamical extinction from $Sr_2Nb_2O_7$ whose space group is $P^{Cmc2_1}_{\bar{1}\,s\,\bar{1}}$. The glide plane (^c_s) is perpendicular to the *b* axis with a glide vector $(\mathbf{c} + \mathbf{a}_4)/2$. The wave number vector of the modulation is $\mathbf{k} = (\frac{1}{2} - \delta)\mathbf{a}^*$. (Since space-group notation $Cmc2_1$ is broadly accepted, the direction of the modulation is taken as the *a* axis.) The reflections $h_10h_3h_4$ with $h_3 + h_4 = 2n + 1$ (n = integer) are kinematically forbidden. Fig. 2.5.3.22 shows a schematic diffraction pattern of $Sr_2Nb_2O_7$ at the [001] incidence. The large

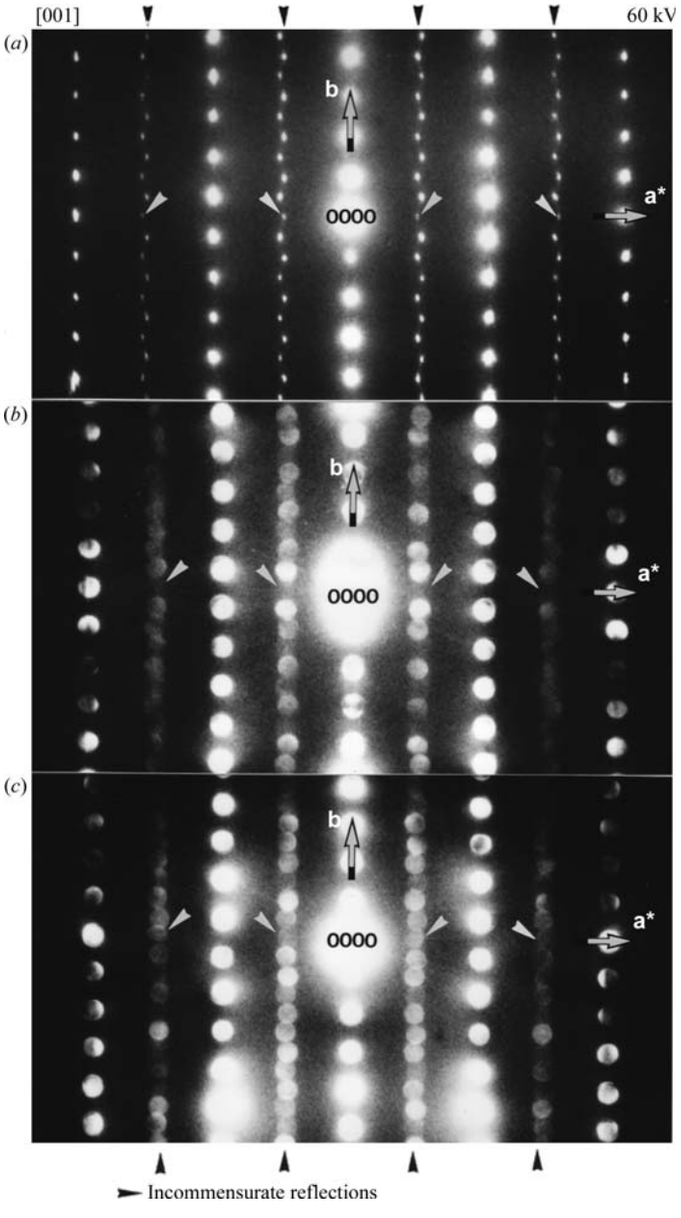


Fig. 2.5.3.23. Diffraction pattern of $Sr_2Nb_2O_7$ taken with [001] incidence at 60 kV. (*a*) Spot diffraction pattern. Kinematically forbidden 0001 and 200$\bar{1}$ incommensurate reflections exhibit definite intensity. (*b*) Zone-axis CBED pattern showing dynamical absence of 0001 and 200$\bar{1}$ incommensurate reflections. (*c*) CBED pattern taken at an incidence with a small tilt from the zone axis to the *b** direction. The kinematically forbidden incommensurate reflections have intensity due to incomplete cancellation of two waves through the *Umweganregung* paths. The wave number vector of modulation is $\mathbf{k} = (\frac{1}{2} - \delta)\mathbf{a}^*$.

and small spots indicate the fundamental (h_4 = 0) and incommensurate ($h_4 \neq 0$) reflections, respectively. *Umweganregung* paths *a* and *b* to the kinematically forbidden 0001 reflection *via* a fundamental reflection in the ZOLZ are drawn.

Fig. 2.5.3.23(*a*) shows a spot diffraction pattern of the incommensurate phase of $Sr_2Nb_2O_7$ taken with the [001] incidence at 60 kV. The incommensurate reflections in which dynamical extinction lines appear at this incidence are those with the indices $h_{1,\text{even}}00h_{4,\text{odd}}$ because $h_3 = 0$ and $h_1 + h_2 = 2n$ due to the lattice type *C* of the average structure.

The reflections in the four columns indicated by black arrowheads are incommensurate reflections. The reflections 0001, 000$\bar{1}$, 200$\bar{1}$ and $\bar{2}$001 designated by white arrowheads are kinematically forbidden but exhibit certain intensities, which are caused by multiple diffraction. Other reflections are fundamental reflections due to the average structure.

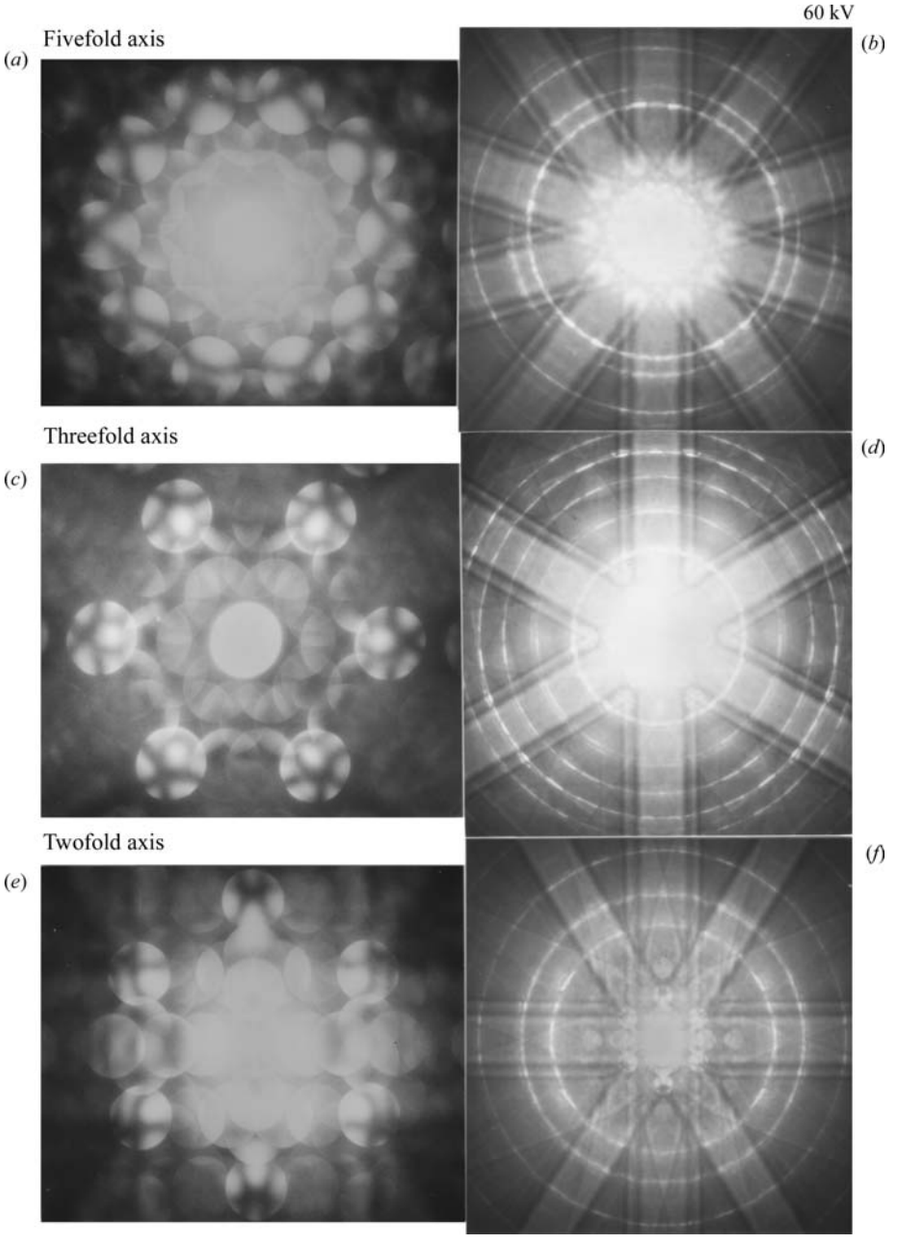


Fig. 2.5.3.24. Three pairs of ZOLZ [(*a*), (*c*) and (*e*)] and HOLZ [(*b*), (*d*) and (*f*)] CBED patterns taken at 60 kV from an area of $Al_{74}Mn_{20}Si_6$ about 3 nm in diameter and about 10 nm thick (Tanaka, Terauchi, Suzuki *et al.*, 1987). Symmetries are (*a*) 10*mm*, (*b*) 5*m*, (*c*) 6*mm*, (*d*) 3*m*, (*e*) 2*mm* and (*f*) 2*mm*.

Fig. 2.5.3.23(*b*) shows a CBED pattern corresponding to Fig. 2.5.3.23(*a*), taken from a specimen area 3 nm in diameter. The excitation errors of two *Umweganregung* paths *a* and *b* are the same at this electron incidence. The reflections 0001, $000\bar{1}$, $200\bar{1}$ and $\bar{2}001$ indicated by white arrowheads show no intensity. Dynamical extinction does not appear as a line in the present case because the width of the extinction line exceeds the disc size of the reflections. Fig. 2.5.3.23(*c*) shows a CBED pattern taken at an incidence slightly tilted toward the b^* axis from that for Fig. 2.5.3.23(*b*) or the [001] zone-axis incidence. The excitation errors are no longer the same for the two *Umweganregung* paths. Thus, it is seen that the kinematically forbidden reflections indicated by white arrowheads have intensities due to incomplete cancellation of waves coming through different paths, which is an additional proof of the dynamical extinction.

2.5.3.5. Symmetry determination of quasicrystals

2.5.3.5.1. Icosahedral quasicrystals

Penrose (1974) demonstrated that a two-dimensional plane can be tiled with thin and fat rhombi to give a pattern with local fivefold rotational symmetries but with no translational symmetry. Mackay (1982) extended the tiling to three dimensions using acute and obtuse rhombohedra, which also resulted in the acquisition of local fivefold rotational symmetries and in a lack of translational symmetry. The three-dimensional space-filling method was later completed by Ogawa (1985). These studies, however, remained a matter of design or geometrical amusement until Shechtman *et al.* (1984) discovered an icosahedral symmetry presumably with long-range structural order in an alloy of Al_6Mn (nominal composition) using electron diffraction. Since then, the term quasicrystalline order, a new class of structural order with no translational symmetry but long-range structural order, has been coined. Levine & Steinhardt (1984) showed that the quasilattice produces sharp diffraction patterns and succeeded in reproducing almost exactly the diffraction pattern obtained by Shechtman *et al*. (1984) using the Fourier transform of a quasi-periodic icosahedral lattice. When analysing X-ray and electron-diffraction data for a quasicrystal, the diffraction peaks can be successfully indexed by six independent vectors pointing to the vertices of an icosahedron. It was then found that the icosahedral quasicrystal can be described in terms of a regular crystal in six

Table 2.5.3.14. *Diffraction groups and CBED symmetries for two icosahedral point groups*

Point group	Diffraction group	BP	WP	DP	±DP
235	$5m_R$	$5m$	5	1	1
					m_R
				m_2	1
	(Projection) $5m1_R$	$10mm$	$5m$	$2 = 1_R$	1
					m_v1_R
				$2m_vm_2$	1
$m\bar{3}\bar{5}$	10_Rmm_R	$10mm$	$5m$	1	2_R
				m_2	2_Rm_2
				m_v	2_Rm_v
	(Projection) $10mm1_R$	$10mm$	$10mm$	2	21_R
				$2m_vm_2$	21_Rm_v

dimensions (*e.g.* Jarić, 1988). A quasicrystal is produced by the intersection of the six-dimensional crystal with an embedded three-dimensional hyperplane (the cut-and-projection technique).

Addition of several per cent of silicon to Al–Mn alloys caused a great increase in the degree of order of the quasicrystal. Bendersky & Kaufman (1986) prepared such a less-strained quasicrystalline $Al_{71}Mn_{23}Si_6$ alloy and determined its point group. They obtained fairly good zone-axis CBED patterns that showed symmetries of $10mm$, $6mm$ and $2mm$ in the ZOLZ discs and $5m$, $3m$ and $2mm$ in HOLZ rings. From these results, they identified the point group to be centrosymmetric $m\bar{3}\bar{5}$. Figs. 2.5.3.24(*a*)–(*f*) show three pairs of CBED patterns taken from an area about 100 nm thick and about 3 nm in diameter of an $Al_{74}Mn_{20}Si_6$ quasicrystal at an accelerating voltage of 60 kV (Tanaka, Terauchi & Sekii, 1987). This quasicrystal was found to have much better ordering than $Al_{71}Mn_{23}Si_6$. The fact that Kikuchi bands are clearly seen in the HOLZ patterns and the profiles of the bands are symmetric with respect to their centre indicates (Figs. 2.5.3.24*b*, *d* and *f*) that the quasicrystal has sufficiently good quality or highly ordered atomic arrangements to perform reliable symmetry determination. Each pair of CBED patterns consists of a ZOLZ pattern and a HOLZ pattern. The former is produced solely by the interaction of ZOLZ reflections, showing distinct symmetries in several discs.

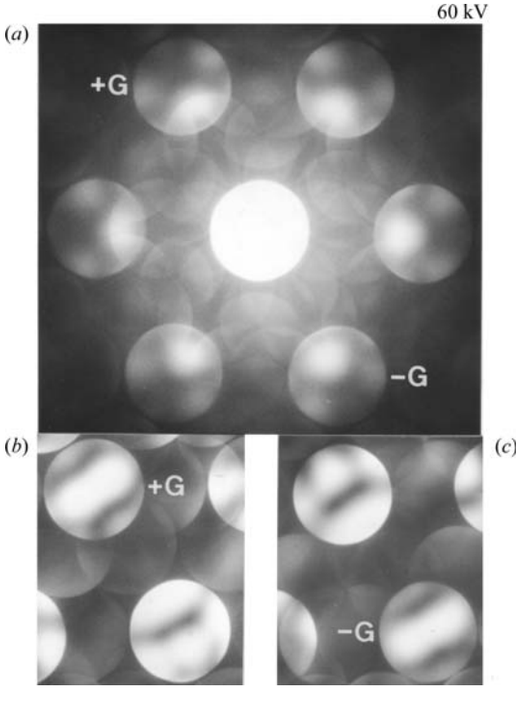


Fig. 2.5.3.25. CBED patterns of $Al_{74}Mn_{20}Si_6$ taken with an electron incidence along the threefold axis. (*a*) Zone-axis pattern showing symmetry $3m$. (*b*,*c*) ±DP showing translational symmetry or 2_R, indicating that the quasicrystal is centrosymmetric.

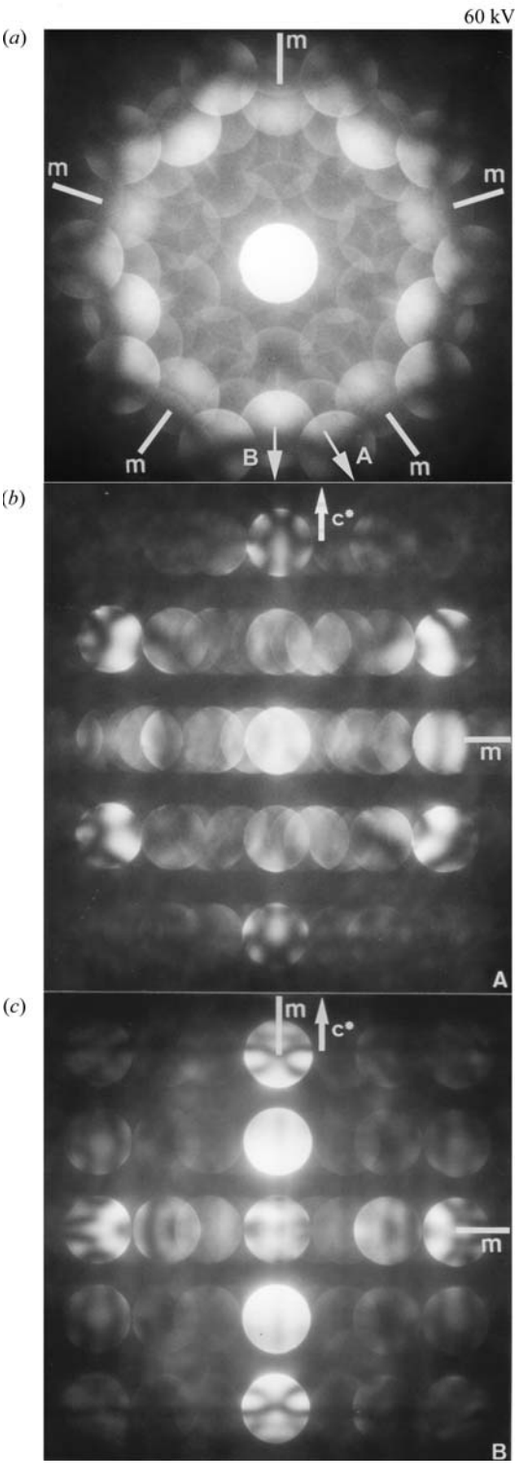
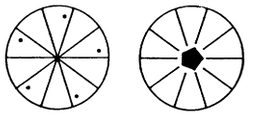
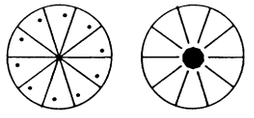
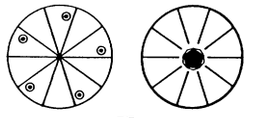
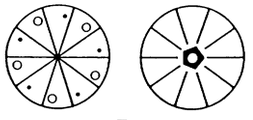
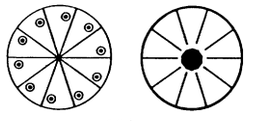
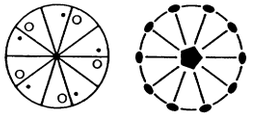
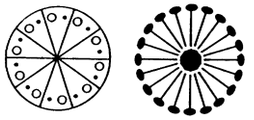
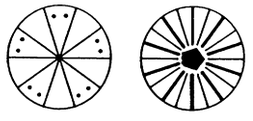
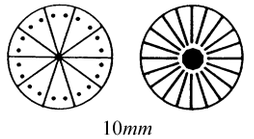
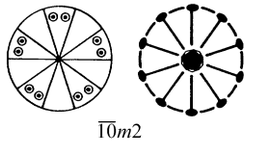
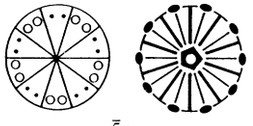
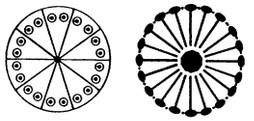


Fig. 2.5.3.26. CBED patterns of metastable $Al_{70}Ni_{15}Fe_{15}$ taken from a 3 nm diameter area. (*a*) Electron incidence along the decagonal axis: symmetry $5m$. (*b*) Electron incidence along direction A indicated in (*a*): symmetry m perpendicular to the decagonal axis. (*c*) Electron incidence along direction B indicated in (*a*): symmetry $2mm$. This alloy is found to be noncentrosymmetric.

The whole pattern of Fig. 2.5.3.24(*a*), formed by ZOLZ reflections, exhibits a tenfold rotation symmetry and two types of mirror symmetry, the resultant symmetry being expressed as

Table 2.5.3.15. *Pentagonal and decagonal point groups constructed by analogy with trigonal and hexagonal point groups*

This table is taken from Saito *et al.* (1992) with the permission of the Japan Society of Applied Physics.

Pentagonal	Decagonal
 5	 10
–	 $\bar{1}0$
 $\bar{5}$	 10/*m*
 5 2	 10 2 2
 5*m*	 10*mm*
–	 $\bar{1}0m2$
 $\bar{5}m$	 10/*mmm*

10*mm*. The whole pattern of Fig. 2.5.3.24(*b*), formed by HOLZ reflections, shows a fivefold rotation symmetry and a type of mirror plane, the resultant symmetry being 5*m*. Figs. 2.5.3.24(*c*) and (*d*) show symmetries 6*mm* and 3*m*, respectively. Figs. 2.5.3.24(*e*) and (*f*) show symmetry 2*mm*. There are two icosahedral point groups, 235 and $m\bar{3}\bar{5}$ (see Table 10.1.4.3 in *IT* A, 2005). The former is noncentrosymmetric with no mirror symmetry but the latter is centrosymmetric. Table 2.5.3.14 shows the diffraction groups expected from these point groups with the incident beam parallel to the fivefold or tenfold axis, and their symmetries appearing in the WP, BP, DP and ±DP. Projection diffraction groups and their symmetries, in which only the interaction between ZOLZ reflections is taken into account, are given in the second row of each pair. Diffraction groups obtained for the other incident-beam directions are omitted because they can be seen in Table 2.5.3.3. The whole-pattern symmetries observed for better-quality images of $Al_{74}Mn_{20}Si_6$ have confirmed the result of Bendersky & Kaufman (1986), *i.e.* the point group $m\bar{3}\bar{5}$. Fig. 2.5.3.25(*a*) shows a zone-axis CBED pattern taken at an electron incidence along the threefold axis. Figs. 2.5.3.25(*b*) and (*c*) show ±DPs taken when tilting the incident beam to excite a low-order strong reflection. The pattern of disc $+G$ agrees with that of disc $-G$ when the former is superposed on the latter with a translation of $-2G$. This symmetry 2_R directly proves that the quasicrystal is centrosymmetric, again confirming the point group as $m\bar{3}\bar{5}$. The lattice type was found to be primitive and no dynamical extinction was observed. Thus, the space group of the alloy was determined to be $Pm\bar{3}\bar{5}$.

Quasicrystals of Al–Mn alloys have been produced by the melt-quenching method and are thermodynamically metastable. Tsai *et al.* (1987) discovered a stable icosahedral phase in $Al_{65}Cu_{20}Fe_{15}$. This alloy has larger grains and is much better quality with less phason strain than $Al_{74}Mn_{20}Si_6$. The discovery of this alloy greatly accelerated the studies of icosahedral quasicrystals. It was found that the lattice type of this phase and of some other Al–Cu–TM (TM = transition metal) alloys is different from that of Al–Mn alloys. That is, Al–Cu–TM alloys display many additional spots in diffraction patterns of twofold rotation symmetry. The patterns were indexed either by all (six) even or all (six) odd, or by a face-centred (*F*) lattice. All the icosahedral quasicrystals known to date belong to the point group $m\bar{3}\bar{5}$; none with the noncentrosymmetric point group 235 have been discovered.

2.5.3.5.2. Decagonal quasicrystals

The first decagonal quasicrystal was found by Bendersky (1985) in an alloy of Al–Mn using the electron-diffraction technique. This phase has periodic order parallel to the tenfold axis, like ordinary crystals, but has quasiperiodic long-range structural order perpendicular to the tenfold axis. The diffraction peaks were indexed by one vector parallel to the tenfold axis and four independent vectors pointing to the vertices of a decagon. Thus, the decagonal quasicrystal is described in terms of a regular crystal in five dimensions.

Two space groups, $P10_5/m$ and $P10_5/mmc$, have been proposed for the alloy by Bendersky (1986) and by Yamamoto & Ishihara (1988), respectively. However, owing to the low quality of the specimens, CBED examination of the alloy could not determine whether the point group is 10/*m* or 10/*mmm*. Furthermore, identification of the space-group symmetry was not possible because observation of dynamical extinction caused by the screw axis and/or the glide plane was difficult. The Al–*M* (*M* = Mn, Fe, Ru, Pt, Pd, ...) quasicrystals found at an early stage were thermodynamically metastable. Subsequently, thermodynamically stable decagonal phases were discovered in the ternary alloys $Al_{65}Cu_{15}Co_{20}$ (Tsai *et al.*, 1989*a*), $Al_{65}Cu_{20}Co_{15}$ (He *et al.*, 1988) and $Al_{70}Ni_{15}Co_{15}$ (Tsai *et al.*, 1989*b*). However, space-group determination was still difficult due to their poor quasicrystallinity.

Tsai *et al.* (1989*c*) succeeded in producing a metastable but good-quality decagonal quasicrystal of $Al_{70}Ni_{15}Fe_{15}$. This alloy was found to be the first decagonal quasicrystal that could tolerate symmetry determination using CBED. The space group was determined to be $P\bar{1}0m2$ by Saito *et al.* (1992).

Fig. 2.5.3.26(*a*) shows a CBED pattern of $Al_{70}Ni_{15}Fe_{15}$ taken with an incidence parallel to the fivefold axis (*c* axis). The pattern clearly exhibits fivefold rotation symmetry and a type of mirror symmetry, the total symmetry being 5*m*. The slowly varying intensity distribution in the discs indicates that the pattern is formed by the interaction between ZOLZ reflections. Thus, the projection approximation should be applied to the analysis of the pattern. Patterns that were related to Fig. 2.5.3.26(*a*) by an inversion were observed when the illuminated specimen area was changed, indicating the existence of inversion domains. Table 2.5.3.15 shows possible pentagonal and decagonal point groups, which are constructed by analogy with the trigonal and hexagonal point groups (Saito *et al.*, 1992).

It can be seen that the point groups that satisfy the observed symmetry 5*m* in the projection approximation are 52, 5*m* and $\bar{1}0m2$. Point group 52 is a possibility because the horizontal

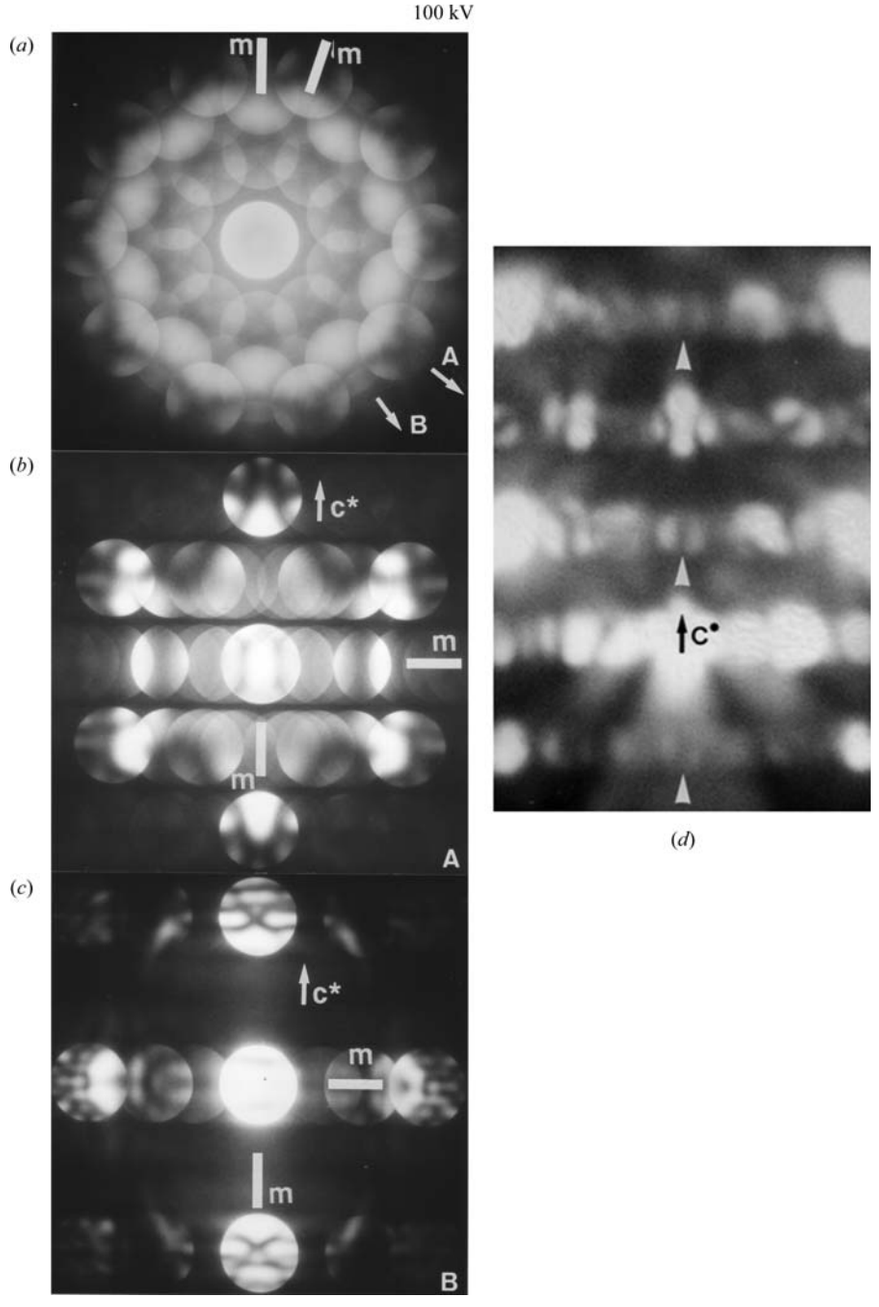


Fig. 2.5.3.27. CBED patterns of metastable $Al_{70}Ni_{20}Fe_{10}$ taken from a 3 nm diameter area. (*a*) Electron incidence along the decagonal axis: symmetry 10*mm*. (*b*) Electron incidence along direction A indicated in (*a*): symmetry 2*mm*. (*c*) Electron incidence along direction B indicated in (*a*): symmetry 2*mm*. (*d*) Reflections 00*l* (*l* = odd) show dynamical extinction lines. This alloy is determined to have the centrosymmetric space group $P10_5/mmc$.

twofold rotation axis is equivalent to the vertical mirror plane in the projection approximation. Figs. 2.5.3.26(*b*) and (*c*) were taken with beam incidences A and B, respectively, as denoted in Fig. 2.5.3.26(*a*). Mirror symmetry perpendicular to the *c* axis is seen in Fig. 2.5.3.26(*b*) and (*c*). Since the mirror symmetry requires a twofold rotation axis or a mirror plane perpendicular to the *c* axis, point groups 52 and $\overline{1}0m2$ remain as possibilities. Fig. 2.5.3.26(*c*) exhibits symmetry 2*mm*. Mirror symmetry parallel to the *c* axis requires the existence of a mirror plane parallel to the axis (a twofold rotation axis is not possible because the fivefold rotation axis already exists.). Since the mirror plane does not exist in point group 52 but does exist in $\overline{1}0m2$, the point group of the alloy is determined to be $\overline{1}0m2$. Examination of the ordinary diffraction patterns of the alloy revealed that the lattice type is primitive with a periodicity of 0.4 nm in the *c* direction and no dynamical extinction was observed. Thus, the space group of $Al_{70}Ni_{15}Fe_{15}$ was determined to be $P\overline{1}0m2$ (Saito *et al.*, 1992) by full use of the potential of CBED. This is the first quasicrystal with a noncentrosymmetric space group. High-resolution electron-microscope images revealed that the quasicrystal is composed of specific pentagonal atom clusters 2 nm in diameter (Tanaka *et al.*, 1993). Dark-field microscopy revealed the existence of inversion domains with an antiphase shift of *c*/2, the polarity being perpendicular to the *c* direction (Tsuda *et al.*, 1993).

Quasicrystals of $Al_{70}Ni_{10+x}Fe_{20-x}$ ($0 \leq x \leq 10$) were investigated by CBED and transmission electron microscopy (Tanaka *et al.*, 1993). The change in space group takes place at $x = 7.5$ upon a sudden decrease of the size of the inversion domains or a rapid mixing of the atom clusters with positive and negative polarities. As a result, the average structure becomes centrosymmetric. A CBED pattern of $Al_{70}Ni_{20}Fe_{10}$ taken at an incidence along the *c* axis shows tenfold rotation symmetry (Fig. 2.5.3.27*a*). CBED patterns taken at incidences A and B (shown in Fig. 2.5.3.27*a*) exhibit two mirror symmetries parallel and perpendicular to the *c* axis (Figs. 2.5.3.27*b* and *c*). Thus, the point group of this phase is

determined to be 10/*mmm*. Fig. 2.5.3.27(*d*) shows a CBED pattern taken by slightly tilting the incident beam to the *c** direction from incidence A. Dynamical extinction lines (arrowheads) are seen in the odd-order reflections along the *c** axis. This indicates the existence of a 10_5 screw axis and a *c*-glide plane. No other reflection absences were observed, implying the lattice type to be primitive. Therefore, the space group of $Al_{70}Ni_{20}Fe_{10}$ is determined to be centrosymmetric $P10_5/mmc$. It was found that the alloys with $0 \leq x \leq 7.5$ belong to the noncentrosymmetric space group $P\overline{1}0m2$ and those with $7.5 < x \leq 15$ belong to the centrosymmetric space group $P10_5/mmc$, keeping the specific polar structure of the basic clusters unchanged.

Another phase was found in the same alloys with $15 < x \leq 17$. This phase showed the same CBED symmetries as the phase with $7.5 < x \leq 15$. The space group of the phase was also determined to be $P10_5/mmc$. However, high-angle annular dark-field (HAADF) observations of the phase with $15 < x \leq 17$ showed that each atom cluster has only one mirror plane of symmetry (Saitoh *et al*., 1997, 1999). This implies that the structure of the specific cluster is changed from that of the phase with $7.5 < x \leq 15$. The clusters are still polar but take ten different orientations, producing centrosymmetric tenfold rotation symmetry on average, which was confirmed by HAADF observations (Saitoh, Tanaka & Tsai, 2001).

These three phases have been found for the similar alloys Al–*M*1–*M*2, where *M*1 = Ni and Cu, and *M*2 = Fe, Co, Rh and Ir (Tanaka *et al.*, 1996). Subsequently, decagonal quasicrystals were found in Al–Pd–Mn, Zn–Mg–RE (RE = Dy, Er, Ho, Lu, Tm and Y) and other alloy systems (Steurer, 2004). There are seven point groups in the decagonal system (Table 2.5.3.15). However, only two point groups, $\overline{1}0m2$ and 10/*mmm*, and two space groups, $P\overline{1}0m2$ and $P10_5/mmc$, are known reliably in real materials to date, though a few other point and space groups have been reported.

For further crystallographic aspects of quasicrystals, the reader is referred to the comprehensive reviews of Tsai (2003) and Steurer (2004), and to a review of more theoretical aspects by Yamamoto (1996).

2.5.4. Electron-diffraction structure analysis (EDSA)[2]

By B. K. Vainshtein and B. B. Zvyagin

2.5.4.1. Introduction

Electron-diffraction structure analysis (EDSA) (Vainshtein, 1964) based on electron diffraction (Pinsker, 1953) is used for the investigation of the atomic structure of matter together with X-ray and neutron diffraction analysis. The peculiarities of EDSA, as compared with X-ray structure analysis, are defined by a strong interaction of electrons with the substance and by a short wavelength λ. According to the Schrödinger equation (see Section 5.2.2) the electrons are scattered by the electrostatic field of an object. The values of the atomic scattering amplitudes, f_e, are three orders higher than those of X-rays, f_x, and neutrons, f_n. Therefore, a very small quantity of a substance is sufficient to obtain a diffraction pattern. EDSA is used for the investigation of very thin single-crystal films, of ~5–50 nm polycrystalline and textured films, and of deposits of finely grained materials and surface layers of bulk specimens. The structures of many ionic crystals, crystal hydrates and hydro-oxides, various inorganic, organic, semiconducting and metallo-organic compounds, of various minerals, especially layer silicates, and of biological structures have been investigated by means of EDSA; it has also been used in the study of polymers, amorphous solids and liquids.

Special areas of EDSA application are: determination of unit cells; establishing orientational and other geometrical relationships between related crystalline phases; phase analysis on the basis of d_{hkl} and I_{hkl} sets; analysis of the distribution of crystallite dimensions in a specimen and inner strains in crystallites as determined from line profiles; investigation of the surface structure of single crystals; structure analysis of crystals, including atomic position determination; precise determination of lattice potential distribution and chemical bonds between atoms; and investigation of crystals of biological origin in combination with electron microscopy (Vainshtein, 1964; Pinsker, 1953; Zvyagin, 1967; Pinsker *et al.*, 1981; Dorset, 1976; Zvyagin *et al.*, 1979).

There are different kinds of electron diffraction (ED) depending on the experimental conditions: high-energy (HEED) (above 30–200 kV), low-energy (LEED) (10–600 V), transmission (THEED) and reflection (RHEED). In electron-diffraction studies use is made of special apparatus – electron-diffraction cameras in which the lens system located between the electron source and the specimen forms the primary electron beam, and the diffracted beams reach the detector without aberration distortions. In this case, high-resolution electron diffraction (HRED) is obtained. ED patterns may also be observed in electron microscopes by a selected-area method (SAD). Other types of electron diffraction are: MBD (microbeam), HDD (high-dispersion), CBD (convergent-beam), SMBD (scanning-beam) and RMBD (rocking-beam) diffraction (see Sections 2.5.2 and 2.5.3). The recent development of electron diffractometry, based on direct intensity registration and measurement by scanning the diffraction pattern against a fixed detector (scintillator followed by photomultiplier), presents a new improved level of EDSA which provides higher precision and reliability of structural data (Avilov *et al*., 1999; Tsipursky & Drits, 1977; Zhukhlistov *et al*., 1997, 1998; Zvyagin *et al.*, 1996).

Electron-diffraction studies of the structure of molecules in vapours and gases is a large special field of research (Vilkov *et al*., 1978). See also *Stereochemical Applications of Gas-Phase Electron Diffraction* (1988).

2.5.4.2. The geometry of ED patterns

In HEED, the electron wavelength λ is about 0.05 Å or less. The Ewald sphere with radius λ^{-1} has a very small curvature and is approximated by a plane. The ED patterns are, therefore, considered as plane cross sections of the reciprocal lattice (RL) passing normal to the incident beam through the point 000, to scale $L\lambda$ (Fig. 2.5.4.1). The basic formula is

$$r = |\mathbf{h}|L\lambda, \text{ or } rd = L\lambda, \qquad (2.5.4.1)$$

where *r* is the distance from the pattern centre to the reflection, **h** is the reciprocal-space vector, *d* is the appropriate interplanar distance and *L* is the specimen-to-screen distance. The deviation of the Ewald sphere from a plane at distance *h* from the origin of the coordinates is $\delta_h = h^2\lambda/2$. Owing to the small values of λ and to the rapid decrease of f_e depending on $(\sin\theta)/\lambda$, the diffracted beams are concentrated in a small angular interval (≤ 0.1 rad).

Single-crystal ED patterns image one plane of the RL. They can be obtained from thin ideal crystalline plates, mosaic single-crystal films or, in the RHEED case, from the faces of bulk single crystals. Point ED patterns can be obtained more easily owing to the following factors: the small size of the crystals (increase in the dimension of RL nodes) and mosaicity – the small spread of crystallite orientations in a specimen (tangential tension of the RL nodes). The crystal system, the parameters of the unit cell and the Laue symmetry are determined from point ED patterns; the probable space group is found from extinctions. Point ED patterns may be used for intensity measurements if the kinematic

[2] Questions related to this section may be addressed to Dr D. L. Dorset (see list of contributing authors).

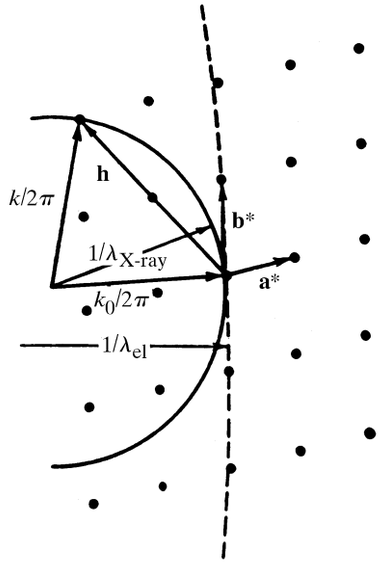


Fig. 2.5.4.1. Ewald spheres in reciprocal space. Dotted line: electrons, solid line: X-rays.

approximation holds true or if the contributions of the dynamic and secondary scattering are not too large.

The indexing of reflections and the unit-cell determination are carried out according to the formulae relating the RL to the DL (direct lattice) (Vainshtein, 1964; Pinsker, 1953; Zvyagin, 1967).

Under electron-diffraction conditions crystals usually show a tendency to lie down on the substrate plane on the most developed face. Let us take this as (001). The vectors **a** and **b** are then parallel, while vector $\mathbf{c}^*$ is normal to this plane, and the RL points are considered as being disposed along direct lines parallel to the axis c^* with constant hk and variable l.

The interpretation of the point patterns as respective RL planes is quite simple in the case of orthogonal lattices. If the lattice is triclinic or monoclinic the pattern of the crystal in the position with the face (001) normal to the incident beam does not have to contain $hk0$ reflections with nonzero h and k because, in general, the planes ab and a^*b^* do not coincide. However, the intersection traces of direct lines hk with the plane normal to them (plane ab) always form a net with periods

$$(a \sin\gamma)^{-1}, (b \sin\gamma)^{-1} \text{ and angle } \gamma' = \pi - \gamma \qquad (2.5.4.2a)$$

(Fig. 2.5.4.2). The points hkl along these directions hk are at distances

$$\eta = ha^* \cos\beta^* + kb^* \cos\alpha^* + lc^* \qquad (2.5.4.3)$$

from the ab plane.

By changing the crystal orientation it is possible to obtain an image of the a^*b^* plane containing $hk0$ reflections, or of other RL planes, with the exception of planes making a small angle with the axis c^*.

In the general case of an arbitrary crystal orientation, the pattern is considered as a plane section of the system of directions hk which makes an angle φ with the plane ab, intersecting it along a direction $[uv]$. It is described by two periods along directions $0h$, $0k$;

$$(a \sin\gamma \cos\psi_h)^{-1}, (b \sin\gamma \cos\psi_k)^{-1}, \qquad (2.5.4.2b)$$

with an angle γ'' between them satisfying the relation

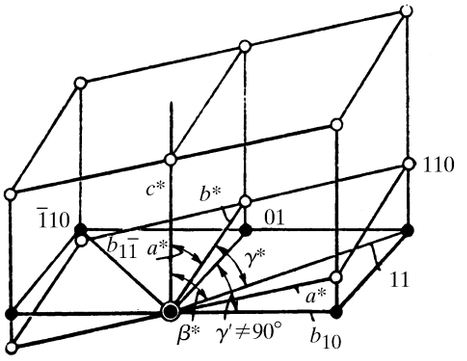


Fig. 2.5.4.2. Triclinic reciprocal lattice. Points: open circles, projection net: black circles.

$$\cos\gamma'' = \sin\psi_h \sin\psi_k - \cos\psi_h \cos\psi_k \cos\gamma, \qquad (2.5.4.2c)$$

and by a system of parallel directions

$$p_h h + p_k k = l; \quad l = 0, \pm1, \pm2, \ldots. \qquad (2.5.4.4)$$

The angles ψ_h, ψ_k are formed by directions $0h$, $0k$ in the plane of the pattern with the plane ab. The coefficients p_h, p_k depend on the unit-cell parameters, angle φ and direction $[uv]$. These relations are used for the indexing of reflections revealed near the integer positions hkl in the pattern and for unit-cell calculations (Vainshtein, 1964; Zvyagin, 1967; Zvyagin *et al.*, 1979).

In RED patterns obtained with an incident beam nearly parallel to the plane ab one can reveal all the RL planes passing through c^* which become normal to the beam at different azimuthal orientations of the crystal.

With the increase of the thickness of crystals (see below, Chapter 5.1) the scattering becomes dynamical and Kikuchi lines and bands appear. Kikuchi ED patterns are used for the estimation of the degree of perfection of the structure of the surface layers of single crystals for specimen orientation in HREM (*IT* C, 2004, Section 4.3.8). Patterns obtained with a convergent beam contain Kossel lines and are used for determining the symmetry of objects under investigation (see Section 2.5.3).

Texture ED patterns are a widely used kind of ED pattern (Pinsker, 1953; Vainshtein, 1964; Zvyagin, 1967). Textured specimens are prepared by substance precipitation on the substrate, from solutions and suspensions, or from gas phase in vacuum. The microcrystals are found to be oriented with a common (developed) face parallel to the substrate, but they have random azimuthal orientations. Correspondingly, the RL also takes random azimuthal orientations, having c^* as the common axis, *i.e.* it is a rotational body of the point RL of a single crystal. Thus, the ED patterns from textures bear a resemblance, from the viewpoint of their geometry, to X-ray rotation patterns, but they are less complicated, since they represent a plane cross section of reciprocal space.

If the crystallites are oriented by the plane (hkl), then the axis $[hkl]^*$ is the texture axis. For the sake of simplicity, let us assume that the basic plane is the plane (001) containing the axes a and b, so that the texture axis is $[001]^*$, *i.e.* the axis c^*. The matrices of appropriate transformations will define a transition to the general case (see *IT* A, 2005). The RL directions hk = constant, parallel to the texture axis, transform to cylindrical surfaces, the points with η_{hkl} = constant are in planes perpendicular to the texture axis, while any 'tilted' lines transform to cones or hyperboloids of rotation. Each point hkl transforms to a ring lying on these surfaces. In practice, owing to a certain spread of c^* axes of single crystals, the rings are blurred into small band sections of a

spherical surface with the centre at the point 000; the oblique cross section of such bands produces reflections in the form of arcs. The main interference curves for texture patterns are ellipses imaging oblique plane cross sections of the cylinders hk (Fig. 2.5.4.3).

At the normal electron-beam incidence (tilting angle $\varphi = 0^\circ$) the ED pattern represents a cross section of cylinders perpendicular to the axis c^*, *i.e.* a system of rings.

On tilting the specimen to an angle φ with respect to its normal position (usually $\varphi \simeq 60^\circ$) the patterns image an oblique cross section of the cylindrical RL, and are called oblique-texture (OT) ED patterns. The ellipses ($hk =$ constant) and layer lines ($l =$ constant) for orthogonal lattices are the main characteristic lines of ED patterns along which the reflections are arranged. The shortcoming of oblique-texture ED patterns is the absence of reflections lying inside the cone formed by rotation of the straight line coming from the point 000 at an angle $(90^\circ - \varphi)$ around the axis c^* and, in particular, of reflections $00l$. However, at $\varphi \lesssim 60$–70° the set of reflections is usually sufficient for structural determination.

For unit-cell determination and reflection indexing the values d (*i.e.* $|\mathbf{h}|$) are used, and the reflection positions defined by the ellipses hk to which they belong and the values η are considered. The periods a^*, b^* are obtained directly from h_{100} and h_{010} values. The period c^*, if it is normal to the plane a^*b^* (γ^* being arbitrary), is calculated as

$$c^* = \eta/l = (h_{hkl}^2 - h_{hk0}^2)^{1/2}/l. \qquad (2.5.4.5a)$$

For oblique-angled lattices

$$c^* = [(h_{l_1+l}^2 + h_{l_1-l}^2 - 2h_l^2)/2]^{1/2}/l. \qquad (2.5.4.5b)$$

In the general case of oblique-angled lattices the coaxial cylinders hk have radii

$$b_{hk} = (1/\sin\gamma)[(h^2/a^2) + (k^2/b^2) - (2hk\cos\gamma/ab)]^{1/2} \qquad (2.5.4.6)$$

and it is always possible to use the measured or calculated values b_{hk} in (2.5.4.5a) instead of h_{hk0}, since

$$\eta = (h_{hkl}^2 - b_{hk}^2)^{1/2}. \qquad (2.5.4.7)$$

In OT patterns the b_{hk} and η values are represented by the lengths of the small axes of the ellipses $B_{hk} = L\lambda b_{hk}$ and the distances of the reflections hkl from the line of small axes (equatorial line of the pattern)

$$D_{hkl} = L\lambda\eta/\sin\varphi = hp + ks + lq. \qquad (2.5.4.8)$$

Analysis of the B_{hk} values gives a, b, γ, while p, s and q are calculated from the D_{hkl} values. It is essential that the components of the normal projections c_n of the axis c on the plane ab measured in the units of a and b are

$$\begin{aligned} x_n &= (c/a)(\cos\beta - \cos\alpha\cos\gamma)/\sin^2\gamma \\ &= -p/q, \\ y_n &= (c/b)(\cos\alpha - \cos\beta\cos\gamma)/\sin^2\gamma \\ &= -s/q. \end{aligned} \qquad (2.5.4.9)$$

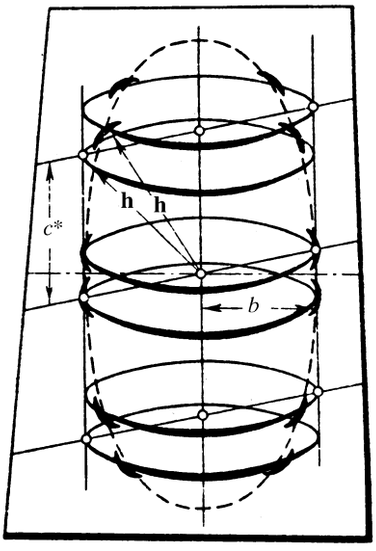


Fig. 2.5.4.3. Formation of ellipses on an electron-diffraction pattern from an oblique texture.

Obtaining x_n, y_n one can calculate

$$c_n = [(x_n a)^2 + (y_n b)^2 + 2x_n y_n ab\cos\gamma]^{1/2}.$$

Since

$$\begin{aligned} d_{001} &= L\lambda/q\sin\varphi, \\ c &= (c_n^2 + d_{001}^2)^{1/2}. \end{aligned} \qquad (2.5.4.10)$$

The α, β values are then defined by the relations

$$\begin{aligned} \cos\alpha &= (x_n a\cos\gamma + y_n b)/c, \\ \cos\beta &= (x_n a + y_n b\cos\gamma)/c. \end{aligned} \qquad (2.5.4.11)$$

Because of the small particle dimensions in textured specimens, the kinematic approximation is more reliable for OT patterns, enabling a more precise calculation of the structure amplitudes from the intensities of reflections.

Polycrystal ED patterns. In this case, the RL is a set of concentric spheres with radii h_{hkl}. The ED pattern, like an X-ray powder pattern, is a set of rings with radii

$$r_{hkl} = h_{hkl}L\lambda. \qquad (2.5.4.12)$$

2.5.4.3. Intensities of diffraction beams

The intensities of scattering by a crystal are determined by the scattering amplitudes of atoms in the crystal, given by (see also Section 5.2.1)

$$\begin{aligned} f_e^{abs}(s) &= 4\pi K\int \varphi(r)r^2\frac{\sin sr}{sr}\,dr; \\ K &= \frac{2\pi me}{h^2};\ f_e = K^{-1}f_e^{abs}, \end{aligned} \qquad (2.5.4.13)$$

where $\varphi(r)$ is the potential of an atom and $s = 4\pi(\sin\theta)/\lambda$. The absolute values of f_e^{abs} have the dimensionality of length L. In EDSA it is convenient to use f_e without K. The dimensionality of

f_e is [potential L^3]. With the expression of f_e in V Å^3 the value K^{-1} in (2.5.4.13) is 47.87 V Å^2.

The scattering atomic amplitudes $f_e(s)$ differ from the respective $f_x(s)$ X-ray values in the following: while $f_x(0) = Z$ (electron shell charge), the atomic amplitude at $s = 0$

$$f_e(0) = 4\pi \int \varphi(r) r^2 \, dr \qquad (2.5.4.14)$$

is the 'full potential' of the atom. On average, $f_e(0) \simeq Z^{1/3}$, but for small atomic numbers Z, owing to the peculiarities in the filling of the electron shells, $f_e(0)$ exhibits within periods of the periodic table of elements 'reverse motion', *i.e.* they decrease with Z increasing (Vainshtein, 1952, 1964). At large $(\sin\theta)/\lambda$, $f_e \simeq Z$. The atomic amplitudes and, consequently, the reflection intensities, are recorded, in practice, up to values of $(\sin\theta)/\lambda \simeq 0.8$–$1.2$ Å^{-1}, *i.e.* up to $d_{\min} \simeq 0.4$–0.6 Å.

The structure amplitude Φ_{hkl} of a crystal is determined by the Fourier integral of the unit-cell potential (see Chapter 1.2),

$$\Phi_{hkl} = \int_{\Omega} \varphi(\mathbf{r}) \exp\{2\pi i(\mathbf{r}\cdot\mathbf{h})\} \, dv_r, \qquad (2.5.4.15)$$

where Ω is the unit-cell volume. The potential of the unit cell can be expressed by the potentials of the atoms of which it is composed:

$$\varphi(\mathbf{r}) = \sum_{\text{cell},\, i} \varphi_{\text{at}\, i}(\mathbf{r} - \mathbf{r}_i). \qquad (2.5.4.16)$$

The thermal motion of atoms in a crystal is taken into account by the convolution of the potential of an atom at rest with the probability function $w(\mathbf{r})$ describing the thermal motion:

$$\varphi_{\text{at}} = \varphi_{\text{at}}(\mathbf{r}) * w(\mathbf{r}). \qquad (2.5.4.17)$$

Accordingly, the atomic temperature factor of the atom in a crystal is

$$f_{eT}[(\sin\theta)/\lambda] = f_e f_T = f_e[(\sin\theta)/\lambda] \exp\{-B[(\sin\theta)/\lambda]^2\}, \qquad (2.5.4.18)$$

where the Debye temperature factor is written for the case of isotropic thermal vibrations. Consequently, the structure amplitude is

$$\Phi_{hkl} = \sum_{\text{cell},\, i} f_{eT_i} \exp\{2\pi i(hx_i + ky_i + lz_i)\}. \qquad (2.5.4.19)$$

This general expression is transformed (see *IT* I, 1952) according to the space group of a given crystal.

To determine the structure amplitudes in EDSA experimentally, one has to use specimens satisfying the kinematic scattering condition, *i.e.* those consisting of extremely thin crystallites. The limit of the applicability of the kinematic approximation (Blackman, 1939; Vainshtein, 1964) can be estimated from the formula

$$A = \lambda \left| \frac{\langle \Phi_{\mathbf{h}} \rangle}{\Omega} \right| t \lesssim 1, \qquad (2.5.4.20)$$

where $\langle \Phi_{\mathbf{h}} \rangle$ is the averaged absolute value of $\Phi_{\mathbf{h}}$ (see also Section 2.5.2). Since $\langle \Phi_{\mathbf{h}} \rangle$ are proportional to $Z^{0.8}$, condition (2.5.4.20) is better fulfilled for crystals with light and medium atoms. Condition (2.5.4.20) is usually satisfied for textured and polycrystalline specimens. But for mosaic single crystals as well, the kinematic approximation limit is, in view of their real structure, substantially wider than estimated by (2.5.4.20) for ideal crystals. The fulfillment of the kinematic law for scattering can be, to a greater or lesser extent, estimated by comparing the decrease of experimental intensity $I_h[(\sin\theta)/\lambda]$ averaged over definite angular intervals, and sums $\sum f^2_{\text{obs}}[(\sin\theta)/\lambda]$ calculated for the same angular intervals.

For mosaic single-crystal films the integral intensity of reflection is

$$I_{\mathbf{h}} = j_0 S \lambda^2 \left| \frac{\Phi_{\mathbf{h}}}{\Omega} \right|^2 \frac{t d_{\mathbf{h}}}{\alpha} \simeq \Phi^2_{\mathbf{h}} d_{\mathbf{h}}; \qquad (2.5.4.21)$$

for textures

$$I_{\mathbf{h}} = j_0 S \lambda^2 \left| \frac{\Phi_{\mathbf{h}}}{\Omega} \right|^2 \frac{tL\lambda p}{2\pi R' \sin\varphi} \simeq \Phi^2_{\mathbf{h}} p / R'. \qquad (2.5.4.22)$$

Here j_0 is the incident electron-beam density, S is the irradiated specimen area, t is the thickness of the specimen, α is the average angular spread of mosaic blocks, R' is the horizontal coordinate of the reflection in the diffraction pattern and p is the multiplicity factor. In the case of polycrystalline specimens the local intensity in the maximum of the ring reflection

$$I_{\mathbf{h}} = j_0 S \lambda^2 \left| \frac{\Phi_{\mathbf{h}}}{\Omega} \right|^2 \frac{t d^2_{\mathbf{h}} p \Delta S}{4\pi L \lambda} \simeq \Phi^2_{\mathbf{h}} d^2_{\mathbf{h}} p \qquad (2.5.4.23)$$

is measured, where ΔS is the measured area of the ring.

The transition from kinematic to dynamic scattering occurs at critical thicknesses of crystals when $A \geq 1$ (2.5.4.20). Mosaic or polycrystalline specimens then result in an uneven contribution of various crystallites to the intensity of the reflections. It is possible to introduce corrections to the experimental structure amplitudes of the first strong reflections most influenced by dynamic scattering by applying in simple cases the two-wave approximation (Blackman, 1939) or by taking into account multibeam theories (Fujimoto, 1959; Cowley, 1981; Avilov *et al.* 1984; see also Chapter 5.2).

The application of kinematic scattering formulae to specimens of thin crystals (5–20 nm) or dynamic corrections to thicker specimens (20–50 nm) permits one to obtain reliability factors between the calculated Φ^{calc} and observed Φ^{obs} structure amplitudes of R = 5–15%, which is sufficient for structural determinations.

With the use of electron-diffractometry techniques, reliability factors as small as R = 2–3% have been reached and more detailed data on the distribution of the inner-crystalline potential field have been obtained, characterizing the state and bonds of atoms, including hydrogen (Zhukhlistov *et al.*, 1997, 1998; Avilov *et al.*, 1999).

The applicability of kinematics formulae becomes poorer in the case of structures with many heavy atoms for which the atomic amplitudes also contain an imaginary component (Shoemaker & Glauber, 1952). The experimental intensity measurement is made by a photo method or by direct recording (Avilov, 1979). In some cases the amplitudes Φ_{hkl} can be determined from dynamic scattering patterns – the bands of equal thickness from a wedge-shaped crystal (Cowley, 1981), or from rocking curves.

2.5.4.4. Structure analysis

The unit cell is defined on the basis of the geometric theory of electron-diffraction patterns, and the space group from extinc-

tions. It is also possible to use the method of converging beams (Section 2.5.3). The structural determination is based on experimental sets of values $|\Phi_{hkl}|^2$ or $|\Phi_{hkl}|$ (Vainshtein, 1964).

The trial-and-error method may be used for the simplest structures. The main method of determination is the construction of the Patterson functions

$$P(xyz) = \frac{1}{\Omega}\left[\Phi_{000}^2 + 2\sum_{hkl=-\infty}^{hkl=+\infty} |\Phi_{hkl}|^2 \cos 2\pi(hx + ky + lz)\right] \tag{2.5.4.24}$$

and their analysis on the basis of heavy-atom methods, superposition methods and so on (see Chapter 2.3). Direct methods are also used (Dorset *et al.*, 1979). Thus the phases of structure factors are calculated and assigned to the observed moduli

$$\Phi_{\mathbf{h}} = |\Phi_{\mathbf{h},\,\mathrm{obs}}| \exp\{i\alpha_{\mathrm{calc}}\}. \tag{2.5.4.25}$$

The distribution of the potential in the unit cell, and, thereby, the arrangement in it of atoms (peaks of the potential) are revealed by the construction of three-dimensional Fourier series of the potential (see also Chapter 1.3)

$$\varphi(xyz) = \frac{1}{\Omega}\sum_{\mathbf{h}} \Phi_{hkl} \exp\{-2\pi i(hx + ky + lz)\} \tag{2.5.4.26a}$$

or projections

$$\varphi'(xy) = \frac{1}{S}\sum_{\mathbf{h}} \Phi_{hk0} \exp\{-2\pi i(hx + ky)\}. \tag{2.5.4.26b}$$

The general formulae (2.5.4.26*a*) and (2.5.4.26*b*) transform, according to known rules, to the expressions for each space group (see *IT* I, 1952). If Φ_{hkl} are expressed in V Å^3 and the volume Ω or the cell area S in Å^3 and Å^2, respectively, then the potential φ is obtained directly in volts, while the projection of the potential φ' is in V Å. The amplitudes $|\Phi_{hkl}|$ are reduced to an absolute scale either according to a group of strong reflections

$$\sum|\Phi_{\mathbf{h}}|^{\mathrm{calc}} = \sum|\Phi_{\mathbf{h}}|^{\mathrm{obs}} \tag{2.5.4.27}$$

or using the Parseval equality

$$\sum_{\mathbf{h}=-\infty}^{+\infty} |\Phi_{\mathbf{h}}|^2 = \Omega^2\langle\varphi^2\rangle = \Omega\sum_{i(\mathrm{cell})} \frac{1}{2\pi^2}\int_0^{\infty} f_{eT_i}^2(s)\, s^2 \,\mathrm{d}s \tag{2.5.4.28}$$

or Wilson's statistical method

$$\langle\Phi^2[(\sin\theta)/\lambda]\rangle = \sum_i f_{eT_i}^2[(\sin\theta)/\lambda]. \tag{2.5.4.29}$$

The term Φ_{000} defines the mean inner potential of a crystal, and is calculated from $f_{\mathrm{e}}(0)$ [(2.5.4.13), (2.5.4.19)]

$$\langle\varphi_{\mathrm{cr}}\rangle = \Phi_{000}/\Omega = \frac{1}{\Omega}\sum f_{\mathrm{e}}(0). \tag{2.5.4.30}$$

The Fourier series of the potential in EDSA possess some peculiarities (Vainshtein, 1954, 1964) which make them different from the electron-density Fourier series in X-ray analysis. Owing to the peculiarities in the behaviour of the atomic amplitudes (2.5.4.13), which decrease more rapidly with increasing $(\sin\theta)/\lambda$ compared with f_{x}, the peaks of the atomic potential

$$\varphi_{\mathrm{at}}(r) = \frac{1}{2\pi^2}\int f_{eT}(s)\frac{\sin sr}{sr} s^2 \,\mathrm{d}s \tag{2.5.4.31}$$

are more 'blurred' and exhibit a larger half-width than the electron-density peaks $\rho_{\mathrm{at}}(r)$. On average, this half-width corresponds to the 'resolution' of an electron-diffraction pattern – about 0.5 Å or better. The potential in the maximum ('peak height') does not depend as strongly on the atomic number as in X-ray analysis:

$$\varphi(0) = \frac{1}{2\pi^2}\int f_{eT}(s)\, s^2 \,\mathrm{d}s \sim Z^{0.75}, \tag{2.5.4.32}$$

while in X-ray diffraction $\rho(0) \sim Z^{1.2}$. In such a way, in EDSA the light atoms are more easily revealed in the presence of heavy atoms than in X-ray diffraction, permitting, in particular, hydrogen atoms to be revealed directly without resorting to difference syntheses as in X-ray diffraction. Typical values of the atomic potential $\varphi(0)$ (which depend on thermal motion) in organic crystals are: H $\sim$ 35, C $\sim$ 165, O 215 V; in Al crystals 330 V, in Cu crystals 750 V.

The EDSA method may be used for crystal structure determination, depending on the types of electron-diffraction patterns, for crystals containing up to several tens of atoms in the unit cell. The accuracy in determination of atomic coordinates in EDSA is about 0.01–0.005 Å on average. The precision of EDSA makes it possible to determine accurately the potential distribution, to investigate atomic ionization, to obtain values for the potential between the atoms and, thereby, to obtain data on the nature of the chemical bond.

If the positions in the cell are occupied only partly, then the measurement of $\varphi_i(0)$ gives information on population percentage.

There is a relationship between the nuclear distribution, electron density and the potential as given by the Poisson equation

$$\nabla^2\varphi(\mathbf{r}) = -4\pi e[\rho_+(\mathbf{r}) - \rho_-(\mathbf{r})]. \tag{2.5.4.33}$$

This makes it possible to interrelate X-ray diffraction, EDSA and neutron-diffraction data. Thus for the atomic amplitudes

$$f_{\mathrm{e}}(s) = 4\pi Ke[Z - f_{\mathrm{x}}(s)]\, s^{-2}, \tag{2.5.4.34}$$

where Z is the nuclear charge and f_{x} the X-ray atomic scattering amplitude, and for structure amplitudes

$$\Phi_{hkl} = \pi Ke[Z_{hkl} - F_{hkl}]|\mathbf{h}|^{-2}, \tag{2.5.4.35}$$

where F_{hkl} is the X-ray structure amplitude of the electron density of a crystal and Z_{hkl} is the amplitude of scattering from charges of nuclei in the cell taking into account their thermal motion. The values Z_{hkl} can be calculated easily from neutron-diffraction data, since the charges of the nuclei are known and the experiment gives the parameters of their thermal motion.

In connection with the development of high-resolution electron-microscopy methods (HREM) it has been found possible to combine the data from direct observations with EDSA methods. However, EDSA permits one to determine the

atomic positions to a greater accuracy, since practically the whole of reciprocal space with 1.0–0.4 Å resolution is used and the three-dimensional arrangement of atoms is calculated. At the same time, in electron microscopy, owing to the peculiarities of electron optics and the necessity for an objective aperture, the image of the atoms in a crystal $\varphi'(\mathbf{x}) * A(\mathbf{x})$ is a convolution, with the aperture function blurring the image up to 1.5–2 Å resolution. In practice, in TEM one obtains only the images of the heaviest atoms of an object. However, the possibility of obtaining a direct image of a structure with all the defects in the atomic arrangement is the undoubted merit of TEM.

2.5.5. Image reconstruction[3]

By B. K. Vainshtein

2.5.5.1. Introduction

In many fields of physical measurements, instrumental and informative techniques, including electron microscopy and computational or analogue methods for processing and transforming signals from objects investigated, find a wide application in obtaining the most accurate structural data. The signal may be radiation from an object, or radiation transmitted through the object, or reflected by it, which is transformed and recorded by a detector.

The image is the two-dimensional signal $I(xy)$ on the observation plane recorded from the whole three-dimensional volume of the object, or from its surface, which provides information on its structure. In an object this information may change owing to transformation of the scattered wave inside an instrument. The real image $J(xy)$ is composed of $I(xy)$ and noise $N(xy)$ from signal disturbances:

$$J(xy) = I(xy) + N(xy). \tag{2.5.5.1}$$

Image-reconstruction methods are aimed at obtaining the most accurate information on the structure of the object; they are subdivided into two types (*Picture Processing and Digital Filtering*, 1975; Rozenfeld, 1969):

(*a*) Image restoration – separation of $I(xy)$ from the image by means of compensation of distortions introduced in it by an image-forming system as well as by an account of the available quantitative data reflecting its structure.

(*b*) Image enhancement – maximum exclusion from the observed image $J(xy)$ (2.5.5.1) of all its imperfections $N(xy)$ from both accidental distortions in objects and various 'noise' in signals and detector, and obtaining $I(xy)$ as the result.

These two methods may be used separately or in combination.

The image should be represented in the form convenient for perception and analysis, *e.g.* in digital form, in lines of equal density, in points of different density, in half-tones or colour form and using, if necessary, a change or reversal of contrast.

Reconstructed images may be used for the three-dimensional reconstruction of the spatial structure of an object, *e.g.* of the density distribution in it (see Section 2.5.6).

This section is connected with an application of the methods of image processing in transmission electron microscopy (TEM). In TEM (see Section 2.5.2), the source-emitted electrons are transmitted through an object and, with the aid of a system of lenses, form a two-dimensional image subject to processing.

Another possibility for obtaining information on the structure of an object is structural analysis with the aid of electron diffraction – EDSA. This method makes use of information in reciprocal space – observation and measurement of electron-diffraction patterns and calculation from them of a two-dimensional projection or three-dimensional structure of an object using the Fourier synthesis. To do this, one has to find the relative phases of the scattered beams.

The wavefunction of an electron-microscopic image is written as

$$\psi_\mathrm{I} = \mathscr{F}^{-1} T \mathscr{F} q \psi_0. \tag{2.5.5.2}$$

Here ψ_0 is the incident plane wave. When the wave is transmitted through an object, it interacts with the electrostatic potential $\varphi(\mathbf{r})$ [$\mathbf{r}(xyz)$ is the three-dimensional vector in the space of the object]; this process is described by the Schrödinger equation (Section 2.5.2.1). As a result, on the exit surface of an object the wave takes the form $q\psi_0(\mathbf{x})$ where q is the transmission function and $\mathbf{x}$ is the two-dimensional vector $\mathbf{x}(xy)$. The diffraction of the wave $q\psi_0$ is described by the two-dimensional Fourier operator:

$$\mathscr{F} q = Q(\mathbf{u}) = \int q(\mathbf{x}) \exp[2\pi i(\mathbf{x}\mathbf{u})]\, \mathrm{d}\mathbf{x}. \tag{2.5.5.3}$$

Here, we assume the initial wave amplitude to be equal to unity and the initial phase to be zero, so that $q\psi_0 = q$, which defines, in this case, the wavefunction in the back focal plane of an objective lens with the reciprocal-space coordinates $\mathbf{u}(u, v)$. The function Q is modified in reciprocal space by the lens transfer function $T(\mathbf{u})$. The scattered wave transformation into an image is described by the inverse Fourier operator $\mathscr{F}^{-1}TQ$.

The process of the diffraction $\mathscr{F} q\psi_0 = Q$, as seen from (2.5.5.1), is the same in both TEM and EDSA. Thus, in TEM under the lens actions $\mathscr{F}^{-1}TQ$ the image formation from a diffraction pattern takes place with an account of the phases, but these phases are modified by the objective-lens transfer function. In EDSA, on the other hand, there is no distorting action of the transfer function and the 'image' is obtained by computing the operation $\mathscr{F}^{-1}Q$.

The computation of projections, images and Fourier transformation is made by discretization of two-dimensional functions on a two-dimensional network of points – pixels in real space $\mathbf{x}(x_j, y_k)$ and in reciprocal space $\mathbf{u}(u_m, v_n)$.

2.5.5.2. Thin weak phase objects at optimal defocus

The intensity distribution $I(xy) \sim |\psi_1|^2$ of an electron wave in the image plane depends not only on the coherent and inelastic scattering, but also on the instrumental functions. The electron wave transmitted through an object interacts with the electrostatic potential $\varphi(\mathbf{r})$ which is produced by the nuclei charges and the electronic shells of the atoms. The scattering and absorption of electrons depend on the structure and thickness of a specimen, and the atomic numbers of the atoms of which it is composed. If an object with the three-dimensional distribution of potential $\varphi(\mathbf{r})$ is sufficiently thin, then the interaction of a plane electron wave ψ_0 with it can be described as the interaction with a two-dimensional distribution of potential projection $\varphi(\mathbf{x})$,

$$\varphi(\mathbf{x}) = \int_0^b \varphi(\mathbf{r})\, \mathrm{d}z, \tag{2.5.5.4}$$

where b is the specimen thickness. It should be noted that, unlike the three-dimensional function of potential $\varphi(\mathbf{r})$ with dimension $[M^{1/2}L^{3/2}T^{-1}]$, the two-dimensional function of potential projection $\varphi(\mathbf{x})$ has the potential-length dimension $[M^{1/2}L^{1/2}T^{-1}]$ which, formally, coincides with the charge dimension. The transmission function, in the general case, has the form

[3] Questions related to this section may be addressed to Dr P. A. Penczek (see list of contributing authors).

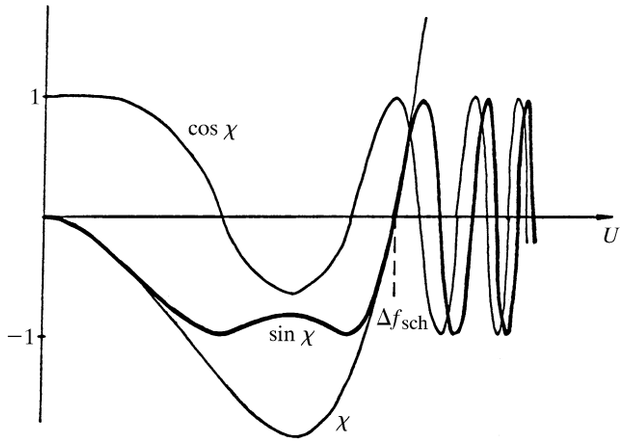


Fig. 2.5.5.1. The χ function and two components of the Scherzer phase function $\sin\chi(U)$ and $\cos\chi(U)$.

$q(\mathbf{x}) = \exp[-i\sigma\varphi(\mathbf{x})]$ (2.5.2.42), and for weak phase objects the approximation $[\sigma\varphi \ll 1]$

$$q(\mathbf{x}) = 1 - i\sigma\varphi(\mathbf{x}) \tag{2.5.5.5}$$

is valid.

In the back focal plane of the objective lens the wave has the form

$$Q(uv)\cdot T(U) \tag{2.5.5.6}$$

$$T = A(U)\exp(i\chi U) \tag{2.5.5.7a}$$

$$\chi(U) = \pi\Delta f\lambda U^2 + \frac{\pi}{2}C_s\lambda^3U^4, \tag{2.5.5.7b}$$

where $U = (u^2 + v^2)^{1/2}$; $\exp[i\chi(U)]$ is the Scherzer phase function (Scherzer, 1949) of an objective lens (Fig. 2.5.5.1), $A(U)$ is the aperture function, C_s the spherical aberration coefficient, and Δf the defocus value [(2.5.2.32)–(2.5.2.35)].

The bright-field image intensity (in object coordinates) is

$$I(xy) = |\psi_1(xy) * t(xy)|^2, \tag{2.5.5.8}$$

where $t = \mathscr{F}^{-1}[T]$. The phase function (2.5.5.7) depends on defocus, and for a weak phase object (Cowley, 1981)

$$I(xy) = 1 + 2\sigma\varphi(xy) * s(xy), \tag{2.5.5.9}$$

where $s = \mathscr{F}^{-1}[A(U)\sin\chi]$, which includes only an imaginary part of function (2.5.5.6). While selecting defocus in such a way that under the Scherzer defocus conditions [(2.5.2.44), (2.5.2.45)] $|\sin\chi| \simeq 1$, one could obtain

$$I(xy) = 1 + 2\sigma\varphi(xy) * a(xy). \tag{2.5.5.10}$$

In this very simple case the image reflects directly the structure of the object – the two-dimensional distribution of the projection of the potential convoluted with the spread function $a = \mathscr{F}^{-1}A$. In this case, no image restoration is necessary. Contrast reversal may be achieved by a change of defocus.

At high resolution, this method enables one to obtain an image of projections of the atomic structure of crystals and defects in the atomic arrangement – vacancies, replacements by foreign atoms, amorphous structures and so on; at resolution worse than atomic one obtains images of dislocations as continuous lines, inserted phases, inclusions *etc.* (Cowley, 1981). It is also possible to obtain images of thin biological crystals, individual molecules, biological macromolecules and their associations.

Image restoration. In the case just considered (2.5.5.10), the projection of potential $\varphi(xy)$, convoluted with the spread function, can be directly observed. In the general case (2.5.5.9), when the aperture becomes larger, the contribution to image formation is made by large values of spatial frequencies U, in which the function $\sin\chi$ oscillates, changing its sign. Naturally, this distorts the image just in the region of appropriate high resolution. However, if one knows the form of the function $\sin\chi$ (2.5.5.7), the true function $\varphi(xy)$ can be restored.

This could be carried out experimentally if one were to place in the back focal plane of an objective lens a zone plate transmitting only one-sign regions of $\sin\chi$ (Hoppe, 1971). In this case, the information on $\varphi(xy)$ is partly lost, but not distorted. To perform such a filtration in an electron microscope is a rather complicated task.

Another method is used (Erickson & Klug, 1971). It consists of a Fourier transformation $\mathscr{F}^{-1}$ of the measured intensity distribution TQ (2.5.5.6) and division of this transform, according to (2.5.5.7*a*,*b*), by the phase function $\sin\chi$. This gives

$$\frac{TQ}{\sin\chi} = Q(uv)A(U). \tag{2.5.5.11a}$$

Then, the new Fourier transformation $\mathscr{F}QA$ yields (in the weak-phase-object approximation) the true distribution

$$\varphi(xy) * a(xy). \tag{2.5.5.11b}$$

The function $\sin\chi$ depending on defocus Δf should be known to perform this procedure. The transfer function can also be found from an electron micrograph (Thon, 1966). It manifests itself in a circular image intensity modulation of an amorphous substrate or, if the specimen is crystalline, in the 'noise' component of the image. The analogue method (optical Fourier transformation for obtaining the image $\sin\chi$) can be used (optical diffraction, see below); digitization and Fourier transformation can also be applied (Hoppe *et al.*, 1973).

The thin crystalline specimen implies that in the back focal objective lens plane the discrete kinematic amplitudes Φ_{hk} are arranged and, by the above method, they are corrected and released from phase distortions introduced by the function $\sin\chi$ (see below) (Unwin & Henderson, 1975).

For the three-dimensional reconstruction (see Section 2.5.6) it is necessary to have the projections of potential of the specimen tilted at different angles α to the beam direction (normal beam incidence corresponds to $\alpha = 0$). In this case, the defocus Δf changes linearly with increase of the distance l of specimen points from the rotation axis $\Delta f_\alpha = \Delta f_0(1 + l\sin\alpha)$. Following the above procedure for passing on to reciprocal space and correction of $\sin\chi$, one can find $\varphi_\alpha(xy)$ (Henderson & Unwin, 1975).

2.5.5.3. An account of absorption

Elastic interaction of an incident wave with a weak phase object is defined on its exit surface by the distribution of potential projection $\varphi(xy)$; however, in the general case, the electron scattering amplitude is a complex one (Glauber & Schomaker, 1953). In such a way, the image itself has the phase and amplitude contrast. This may be taken into account if one considers not only the potential projection $\varphi(xy)$, but also the 'imaginary potential'

$\mu(xy)$ which describes phenomenologically the absorption in thin specimens. Then, instead of (2.5.5.5), the wave on the exit surface of a specimen can be written as

$$q(xy) = 1 - i\sigma\varphi(xy) - \mu(xy) \tag{2.5.5.12}$$

and in the back focal plane if $\Phi = \mathscr{F}\varphi$ and $M = \mathscr{F}\mu$

$$Q(uv) = \delta(uv) - i\sigma\Phi(uv) - M(uv). \tag{2.5.5.13}$$

Usually, μ is small, but it can, nevertheless, make a certain contribution to an image. In a sufficiently good linear approximation, it may be assumed that the real part $\cos\chi$ of the phase function (2.5.5.7*a*) affects $M(uv)$, while $\Phi(xy)$, as we know, is under the action of the imaginary part $\sin\chi$.

Thus, instead of (2.5.5.6), one can write

$$Q(\exp i\chi) = \delta(\mathbf{u}) - i\sigma\Phi(\mathbf{u})\sin\chi - M(\mathbf{u})\cos\chi, \tag{2.5.5.14}$$

and as the result, instead of (2.5.5.10),

$$\begin{aligned} I(xy) = 1 &+ 2\sigma\varphi(xy) * \mathscr{F}^{-1}(\sin\chi) * a(U) \\ &- 2\mu(xy) * \mathscr{F}^{-1}(\cos\chi) * a(U). \end{aligned} \tag{2.5.5.15}$$

The functions $\varphi(xy)$ and $\mu(xy)$ can be separated by object imaging using the through-focus series method. In this case, using the Fourier transformation, one passes from the intensity distribution (2.5.5.15) in real space to reciprocal space. Now, at two different defocus values Δf_1 and Δf_2 [(2.5.5.6), (2.5.5.7*a*,*b*)] the values $\Phi(\mathbf{u})$ and $M(\mathbf{u})$ can be found from the two linear equations (2.5.5.14). Using the inverse Fourier transformation, one can pass on again to real space which gives $\varphi(\mathbf{x})$ and $\mu(\mathbf{x})$ (Schiske, 1968). In practice, it is possible to use several through-focus series and to solve a set of equations by the least-squares method.

Another method for processing takes into account the simultaneous presence of noise $N(\mathbf{x})$ and transfer function zeros (Kirkland *et al.*, 1980). In this method the space frequencies corresponding to small values of the transfer function modulus are suppressed, while the regions where such a modulus is large are found to be reinforced.

2.5.5.4. Thick crystals

When the specimen thickness exceeds a certain critical value (~50–100 Å), the kinematic approximation does not hold true and the scattering is dynamic. This means that on the exit surface of a specimen the wave is not defined as yet by the projection of potential $\varphi(xy) = \int\varphi(\mathbf{r})\,\mathrm{d}z$ (2.5.5.3), but one has to take into account the interaction of the incident wave ψ_0 and of all the secondary waves arising in the whole volume of a specimen.

The dynamic scattering calculation can be made by various methods. One is the multislice (or phase-grating) method based on a recurrent application of formulae (2.5.5.3) for n thin layers Δz_i thick, and successive construction of the transmission functions q_i (2.5.5.4), phase functions $Q_i = \mathscr{F}q_i$, and propagation function $p_k = [k/2\pi i\Delta z]\exp[ik(x^2 + y^2)/2\Delta z]$ (Cowley & Moodie, 1957).

Another method – the scattering matrix method – is based on the solution of equations of the dynamic theory (Chapter 5.2). The emerging wave on the exit surface of a crystal is then found to diffract and experience the transfer function action [(2.5.5.6), (2.5.5.7*a*,*b*)].

The dynamic scattering in crystals may be interpreted using Bloch waves:

$$\Psi^j(\mathbf{r}) = \sum_H C_H^j \exp(-2\pi i\mathbf{k}_H^j \cdot \mathbf{r}). \tag{2.5.5.16}$$

It turns out that only a few (bound and valence Bloch waves) have strong excitation amplitudes. Depending on the thickness of a crystal, only one of these waves or their linear combinations (Kambe, 1982) emerges on the exit surface. An electron-microscopic image can be interpreted, at certain thicknesses, as an image of one of these waves [with a correction for the transfer function action (2.5.5.6), (2.5.5.7*a*,*b*)]; in this case, the identical images repeat with increasing thickness, while, at a certain thickness, the contrast reversal can be observed. Only the first Bloch wave which arises at small thickness, and also repeats with increasing thickness, corresponds to the projection of potential $\varphi(xy)$, *i.e.* the atom projection distribution in a thin crystal layer.

An image of other Bloch waves is defined by the function $\varphi(\mathbf{r})$, but their maxima or minima do not coincide, in the general case, with the atomic positions and cannot be interpreted as the projection of potential. It is difficult to reconstruct $\varphi(xy)$ from these images, especially when the crystal is not ideal and contains imperfections. In these cases one resorts to computer modelling of images at different thicknesses and defocus values, and to comparison with an experimentally observed pattern.

The imaging can be performed directly in an electron microscope not by a photo plate, but using fast-response detectors with digitized intensity output online. The computer contains the necessary algorithms for Fourier transformation, image calculation, transfer function computing, averaging, and correction for the observed and calculated data. This makes possible the interpretation of the pattern observed directly in experiment (Herrmann *et al.*, 1980).

2.5.5.5. Image enhancement

The real electron-microscope image is subdivided into two components:

$$J(xy) = I(xy) + N(xy). \tag{2.5.5.17}$$

The main of these, $I(xy)$, is a two-dimensional image of the 'ideal' object obtained in an electron microscope with instrumental functions inherent to it. However, in the process of object imaging and transfer of this information to the detector there are various sources of noise. In an electron microscope, these arise owing to emission-current and accelerating-voltage fluctuations, lens-supplying current (temporal fluctuations), or mechanical instabilities in a device, specimen or detector (spatial shifts). The two-dimensional detector (*e.g.* a photographic plate) has structural inhomogeneities affecting a response to the signal. In addition, the specimen is also unstable; during preparation or imaging it may change owing to chemical or some other transformations in its structure, thermal effects and so on. Biological specimens scatter electrons very weakly and their natural state is moist, while in the electron-microscope column they are under vacuum conditions. The methods of staining (negative or positive), *e.g.* of introducing into specimens substances containing heavy atoms, as well as the freeze-etching method, somewhat distort the structure of a specimen. Another source of structure perturbation is radiation damage, which can be eliminated at small radiation doses or by using the cryogenic technique. The structure of stained specimens is affected by stain graininess. We assume that all the deviations $\Delta I_k(xy)$ of a specimen image from the 'ideal' image $I_k(xy)$ are included in the noise term $N_k(xy)$. The substrate may also be inhomogeneous. All kinds of perturbations cannot be separated and they appear on an electron microscope image as the full noise content $N(xy)$.

The image enhancement involves maximum noise suppression $N(xy)$ and hence the most accurate separation of a useful signal $I(xy)$ from the real image $J(xy)$ (2.5.5.1). At the signal/noise ratio $I/N \simeq 1$ such a separation appears to be rather complicated. But in some cases the real image reflects the structure sufficiently well, *e.g.* during the atomic structure imaging of some crystals ($I/N > 10$). In other cases, especially of biological specimen imaging, the noise N distorts substantially the image, (I/N) ~ 5–10. Here one should use the methods of enhancement. This problem is usually solved by the methods of statistical processing of sets of images J_k ($k = 1, \ldots, n$). If one assumes that the informative signal $I_k(xy)$ is always the same, then the noise error $N(xy)$ may be reduced.

The image enhancement methods are subdivided into two classes:

(a) image averaging in real space xy;

(b) Fourier analysis and filtration in reciprocal space.

These methods can be used separately or in combination. The enhancement can be applied to both the original and the restored images; there are also methods of simultaneous restoration and enhancement.

The image can be enhanced by analogue (mainly optical and photographic) methods or by computational methods for processing digitized functions in real and reciprocal space.

The cases where the image has translational symmetry, rotational symmetry, and where the image is asymmetric will be considered.

Periodic images. An image of the crystal structure with atomic or molecular resolution may be brought to self-alignment by a shift by a and b periods in a structure projection. This can be performed photographically by printing the shifted image on the same photographic paper or, *vice versa*, by shifting the paper (McLachlan, 1958).

The Fourier filtration method for a periodic image I_p with noise N is based on the fact that in Fourier space the components $\mathscr{F}I_p$ and $\mathscr{F}N$ are separated. Let us carry out the Fourier transformation of the periodic signal I_p with the periods a, b and noise N:

$$\begin{aligned}\mathscr{F}J &= \mathscr{F}[I_p(xy) + N(xy)] \\ &= \textstyle\int I_p(xy) \exp[2\pi i(hx + ky)]\, \mathrm{d}x\, \mathrm{d}y + \mathscr{F}N \\ &= \textstyle\sum \Phi_{hk} \delta(\mathbf{u} - \mathbf{u}_{hk}) + \mathscr{F}N; \qquad (2.5.5.18) \\ \mathbf{u}_{hk} &= h\mathbf{a}^* + k\mathbf{b}^*.\end{aligned}$$

The left part of (2.5.5.18) represents the Fourier coefficients Φ_{hk} distributed discretely with periods a^* and b^* in the plane $\mathbf{u}(uv)$. This is the two-dimensional reciprocal lattice. The right-hand side of (2.5.5.18) is the Fourier transform $\mathscr{F}N$ distributed continuously in the plane. Thus these parts are separated. Let us 'cut out' from distribution (2.5.5.18) only Φ_{hk} values using the 'window' function $w(uv)$. The window should match each of the real peaks Φ_{hk} which, owing to the finite dimensions of the initial periodic image, are not points, as this is written in an idealized form in (2.5.5.18) with the aid of δ functions. In reality, the 'windows' may be squares of about $a^*/10$, $b^*/10$ in size, or a circle. Performing the Fourier transformation of product (2.5.5.18) without $\mathscr{F}N$, and set of windows $w(\mathbf{u}) = w(uv) * \sum_{h,k} \delta(\mathbf{u} - h\mathbf{a}^* - k\mathbf{b}^*)$, we obtain:

$$\begin{aligned}J(xy) &= \mathscr{F}^{-1}\{w(\mathbf{u}) \sum_{h,k} \Phi_{h,k} \delta(\mathbf{u} - \mathbf{u}_{h,k})\} \\ &= W(xy) * I_p(\mathbf{x}), \qquad (2.5.5.19)\end{aligned}$$

the periodic component without the background, $W(xy) = \mathscr{F}^{-1}w(\mathbf{u})$. The zero coefficient Φ_{00} in (2.5.5.19) should be decreased, since it is due, in part, to the noise. When the window w is sufficiently small, I_p in (2.5.5.19) represents the periodic distribution $\langle I \rangle$ (average over all the unit cells of the projection) included in I_p (2.5.5.18). Nevertheless, some error from noise in an image does exist, since with Φ_{hk} we also introduced into the inverse Fourier transformation the background transform values $\mathscr{F}^{-1}N_{hk}$ which are within the 'windows'.

This approach is realized by an analogue method [optical diffraction and filtering of electron micrographs in a laser beam (Klug & Berger, 1964)] and can also be carried out by computing.

As an example, Fig. 2.5.5.2(b) shows an electron micrograph of the periodic structure of a two-dimensional protein crystal, while Fig. 2.5.5.2(c) represents optical diffraction from this layer. In order to dissect the aperiodic component $\mathscr{F}N$ in a diffraction plane, according to the scheme in Fig. 2.5.5.2(a), one places a mask with windows covering reciprocal-lattice points. After such a filtration, only the I_p component makes a contribution during the image formation by means of a lens, while the component $\mathscr{F}N$ diffracted by the background is delayed. As a result, an optical pattern of the periodic structure is obtained (Fig. 2.5.5.2d).

Optical diffractometry also assists in determining the parameters of a two-dimensional lattice and its symmetry.

Using the same method, one can separate the superimposed images of two-dimensional structures with different periodicity and in different orientation, the images of the 'near' and 'far' sides of tubular periodic structures with monomolecular walls (Klug & DeRosier, 1966; Kiselev *et al.*, 1971), and so on.

Computer filtering involves measuring the image optical density J_{obs}, digitization, and Fourier transformation (Crowther & Amos, 1971). The sampling distance usually corresponds to one-third of the image resolution. When periodic weak phase objects are investigated, the transformation (2.5.5.18) yields the Fourier coefficients. If necessary, we can immediately make corrections in them using the microscope transfer function according to (2.5.5.6), (2.5.5.7a,b) and (2.5.5.11a), and thereby obtain the true kinematic amplitudes Φ_{hk}. The inverse transformation (2.5.5.16) gives a projection of the structure (Unwin & Henderson, 1975; Henderson & Unwin, 1975).

Sometimes, an observed image $J(\mathbf{x})$ is 'noised' by the $N(\mathbf{x})$ to a great extent. Then, one may combine data on real and reciprocal space to construct a sufficiently accurate image. In this case, the electron-diffraction pattern is measured and structure-factor moduli from diffraction reflection intensities $I_{hk,\,obs}$ are obtained:

$$|\Phi_{hk,\,obs}| \sim \sqrt{I_{hk,\,obs}}. \qquad (2.5.5.20)$$

At the same time, the structure factors

$$\Phi_{hk,\,calc} = |\Phi_{hk,\,calc}| \exp(i\alpha_{hk,\,calc}) \qquad (2.5.5.21)$$

are calculated from the processed structure projection image by means of the Fourier transformation. However, owing to poor image quality we take from these data only the values of phases α_{hk} since they are less sensitive to scattering density distortions than the moduli, and construct the Fourier synthesis

$$\begin{aligned}I(xy) = \sum_{hk} &|\Phi_{hk,\,obs}| \exp(i\alpha_{hk,\,calc}) \\ &\times \exp[2\pi i(hx + ky)]. \qquad (2.5.5.22)\end{aligned}$$

Here the possibilities of combining various methods open up, *e.g.* for obtaining the structure-factor moduli from X-ray

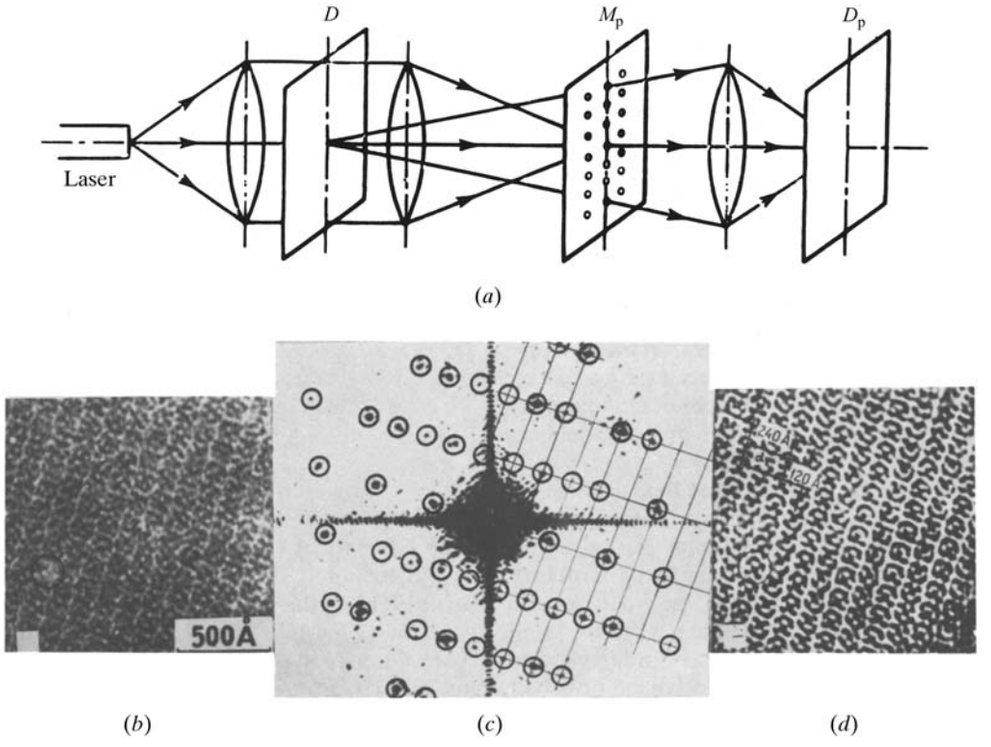


Fig. 2.5.5.2. (*a*) Diagram of an optical diffractometer. *D* is the object (an electron micrograph), M_p is the diffraction plane and a mask that transmits only Φ_{hk}, D_p is the plane of the (filtered) image; (*b*) an electron micrograph of a crystalline layer of the protein phosphorylase *b*; (*c*) its optical diffraction pattern (the circles correspond to the windows in the mask that transmits only the Φ_{hk} diffracted beams from the periodic component of the image); (*d*) the filtered image. Parts (*b*)–(*d*) are based on the article by Kiselev *et al.* (1971).

diffraction, and phases from electron microscopy, and so on (Gurskaya *et al.*, 1971).

Images with point symmetry. If a projection of an object (and consequently, the object itself) has a rotational *N*-fold axis of symmetry, the structure coincides with itself on rotation through the angle $2\pi/N$. If the image is rotated through arbitrary angles and is aligned photographically with the initial image, then the best density coincidence will take place at a rotation through $\alpha = (k2\pi/N)$ $(k = 1, \ldots, N)$ which defines *N*. The pattern averaging over all the rotations will give the enhanced structure image with an $(N)^{1/2}$ times reduced background (Markham *et al.*, 1963).

Rotational filtering can be performed on the basis of the Fourier expansion of an image in polar coordinates over the angles (Crowther & Amos, 1971).

$$I(r, \psi) = \sum_{n=-\infty}^{+\infty} g_n(r) \exp(in\varphi). \qquad (2.5.5.23)$$

The integral over the radius from azimuthal components g_n gives their power

$$p_n \sim \int_0^a |g_n|^2 r \, \mathrm{d}r, \qquad (2.5.5.24)$$

where *a* is the maximum radius of the particle. A set p_n forms a spectrum, the least common multiple *N* of strong peaks defining the *N*-fold symmetry. The two-dimensional reconstructed image of a particle with rotational symmetry is defined by the synthesis (2.5.5.24) with $n = 0, N, 2N, 3N$.

Asymmetric images. In this case, a set of images is processed by computational or analogue methods. The initial selection of images involves the fulfilment of the maximum similarity condition.

The averaging of *n* images in real space gives

$$I_{\text{enh}} = (1/n) \sum_{k=1}^{n} J_k(xy) = \langle I_k \rangle(xy) + (1/n) \sum N_k(xy). \qquad (2.5.5.25)$$

The signal/noise ratio on an average image is $(n)^{1/2}$ times enhanced.

The degree of similarity and accuracy of superposition of two images with an account both of translational and angular shifts is estimated by a cross-correlation function[4] of two selected images J_1 and J_2 (Frank, 1975, 1980).

$$\begin{aligned} k(\mathbf{x}') &= J_1 * J_2 = \int J_1(x) J_2(x + x') \, \mathrm{d}x \\ &= k_{I_1 I_2} + k_{I_1 N_2} + k_{I_2 N_1} + k_{N_1 N_2}. \end{aligned} \qquad (2.5.5.26)$$

The value $k(\mathbf{x}')$ is the measure of image similarity, the x' coordinate of the maximum indicates the shift of the images relative to each other. The first term of the resultant expression (2.5.5.26) is the cross-correlation function of noise-corrected images being compared, the second and third terms are approximately equal to zero, since the noise does not correlate with the signal; the last term is the autocorrelation function of the noise (Cramér, 1954; Frank, 1975, 1980).

The calculation of a correlation function is performed by means of Fourier transformation on the basis of the convolution theorem, since the Fourier transformation of the product of the Fourier transform of function J_1 and the conjugated Fourier transform function J_2 gives the cross-correlation function of the initial functions:

[4] At $I_j = I_k$ this is the autocorrelation function, an analogue of the Patterson function used in crystallography.

$$k = \mathscr{F}^{-1}[\mathscr{F}J_1 \cdot \mathscr{F}^* J_2]. \tag{2.5.5.27}$$

The probability density of samples for images has the form

$$p(J_1 J_2 \ldots J_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} \times \exp\left[\frac{-1}{2\sigma^2}\sum^{n}\int [J_k(\mathbf{x}+\mathbf{x}_k) - J(\mathbf{x})]^2 \, \mathrm{d}x\right]. \tag{2.5.5.28}$$

Here J is the tentative image (as such, a certain 'best' image can first be selected, while at the repeated cycle an average image is obtained), $J_k(\mathbf{x})$ is the image investigated, σ is the standard deviation of the normal distribution of noises and x_k the relative shift of the image. This function is called a likelihood function; it has maxima relative to the parameters $J(x)$, x_k, σ. The average image and dispersion are

$$J(\mathbf{x}) = (1/n)\sum^{n}[J_k(\mathbf{x}-\mathbf{x}_k)],$$
$$\sigma^2 = (1/n)\sum^{n}[J_k(\mathbf{x}-\mathbf{x}_k) - J(\mathbf{x})]^2. \tag{2.5.5.29}$$

This method is called the maximum-likelihood method (Cramér, 1954; Kosykh *et al.*, 1983).

It is convenient to carry out the image alignment, in turn, with respect to translational and angular coordinates. If we start with an angular alignment we first use autocorrelation functions or power spectra, which have the maximum and the symmetry centre at the origin of the coordinates. The angular correlation maximum

$$f(\theta') = \int f_k(\theta - \theta') f_e(\theta) \, \mathrm{d}\theta \tag{2.5.5.30}$$

gives the mutual angle of rotation of two images.

Then we carry out the translational alignment of rotationally aligned images using the translational correlation function (2.5.5.26) (Langer *et al.*, 1970).

In the iteration alignment method, the images are first translationally aligned and then an angular shift is determined in image space in polar coordinates with the centre at the point of the best translational alignment. After the angular alignment the whole procedure may be repeated (Steinkilberg & Schramm, 1980).

The average image obtained may have false high-frequency components. They can be excluded by multiplying its Fourier components by some function and suppressing high-space frequencies, for instance by an 'artificial temperature factor' $\exp\{-B|\mathbf{u}|^2\}$.

For a set of similar images the Fourier filtration method can also be used (Ottensmeyer *et al.*, 1977). To do this, one should prepare from these images an artificial 'two-dimensional crystal', *i.e.* place them in the same orientation at the points of the two-dimensional lattice with periods a, b.

$$J = \sum_{k=1}^{n} J_k(\mathbf{x} - \mathbf{t}_p); \ \mathbf{t} = p_1\mathbf{a} + p_2\mathbf{b}. \tag{2.5.5.31}$$

The processing is then performed according to (2.5.5.18), (2.5.5.19); as a result one obtains $\langle I(xy)\rangle$ with reduced background. Some translational and angular errors in the arrangement of the images at the artificial lattice points act as an artificial temperature factor. The method can be realized by computing or by optical diffraction.

2.5.6. Three-dimensional reconstruction[5]

By B. K. Vainshtein and P. A. Penczek

2.5.6.1. The object and its projection

In electron microscopy (EM) and single-particle reconstruction, three-dimensional (3D) reconstruction methods are used for studying biological structures; that is, symmetric or asymmetric associations of biomacromolecules (muscles, spherical and rod-like viruses, bacteriophages, individual proteins and ribosomes) (Frank, 2006). The electron microscope is used to obtain parallel-beam two-dimensional (2D) projections $\varphi_2(\mathbf{x}, \boldsymbol{\tau})$ (ray transform) of frozen hydrated 3D macromolecules $\varphi_3(\mathbf{r})$ suspended in random orientations (Fig. 2.5.6.1). The function $\varphi_2(\mathbf{x}_\tau)$ is the 2D projection of the 3D molecular electron distribution $\varphi_3(\mathbf{r})$. One can also consider one-dimensional (1D) projections $\varphi_1(s, \boldsymbol{\tau})$ of multidimensional distributions; the set of these projections is called a Radon transform. For 2D distributions, ray and Radon transforms differ only in the notation. For ($d > 2$)-dimensional distributions the two are different: in a ray transform the integrals are calculated over straight lines and yield ($d - 1$)-dimensional projections, while in Radon transforms the integrals are calculated over ($d - 1$)-dimensional hyperplanes and yield 1D projections. In electron microscopy, Radon transforms are not directly measurable, but can be formed computationally and used in intermediate steps of the 3D reconstruction or in alignment procedures (Radermacher, 1994).

Within the linear weak-phase-object approximation of the image-formation process in the microscope [see equation (2.5.2.42) in Section 2.5.2 of this chapter], 2D projections represent line integrals of the potential of the particle under examination convoluted with the point-spread function of the microscope, s, so, using (2.5.2.43),

$$I(xy) = 1 + 2\sigma s(xy) * \varphi(xy) = 1 + 2\sigma s(xy) * \int \varphi_3(\mathbf{r}) \, \mathrm{d}z. \tag{2.5.6.1}$$

Since

$$\varphi_2(\mathbf{x}_\tau) = \int \varphi_3(\mathbf{r}) \, \mathrm{d}\tau, \quad \boldsymbol{\tau} \perp \mathbf{x}, \tag{2.5.6.2}$$

we have

$$I(xy) = 1 + 2\sigma s(\mathbf{x}_\tau) * \varphi_2(\mathbf{x}_\tau), \quad \boldsymbol{\tau} \perp \mathbf{x}. \tag{2.5.6.3}$$

If we omit constant terms, we obtain

$$I(xy) = s(\mathbf{x}_\tau) * \varphi_2(\mathbf{x}_\tau) = s(\mathbf{x}_\tau) * \int \varphi_3(\mathbf{r}) \, \mathrm{d}\tau, \quad \boldsymbol{\tau} \perp \mathbf{x}. \tag{2.5.6.4}$$

In this section, we will assume that all images were collected using the same defocus setting, so the point-spread function s is constant and does not depend on the projection direction $\boldsymbol{\tau}$. Thus, we will concern ourselves with the inversion of the projection problem

$$\varphi_2(\mathbf{x}_\tau) = \int \varphi_3(\mathbf{r}) \, \mathrm{d}\tau, \quad \boldsymbol{\tau} \perp \mathbf{x}. \tag{2.5.6.5}$$

[5] The original version of Section 2.5.6, written by the late B. K. Vainshtein, is here updated and expanded by P. A. Penczek.

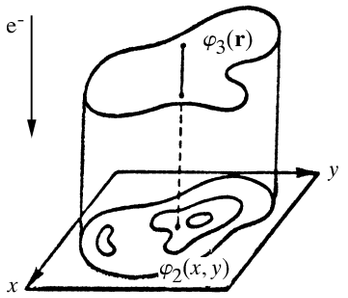


Fig. 2.5.6.1. A three-dimensional object φ_3 and its two-dimensional projection φ_2. The electron beam penetrates the specimen in the direction of the z axis.

The projection direction is defined by a unit vector $\boldsymbol{\tau}(\theta, \psi)$ and it is formed on the plane $\mathbf{x}$ perpendicular to $\boldsymbol{\tau}$. The set of various projections $\varphi_2(\mathbf{x}_{\tau_i}) = \varphi_{2_i}(\mathbf{x}_i)$ may be assigned by a discrete or continuous set of points on a unit sphere $|\tau| = 1$ (Fig. 2.5.6.2).

In Fourier space, the relation between an object and its projection is referred to as the central section theorem: the Fourier transformation of projection φ_2 of a 3D object φ_3 is the central (*i.e.*, passing though the origin of reciprocal space) 2D plane cross section of a 3D transform perpendicular to the projection vector (Bracewell, 1956; DeRosier & Klug, 1968; Crowther, Amos *et al.*, 1970). In Cartesian coordinates, a 3D Fourier transform is

$$\begin{aligned}\mathscr{F}[\varphi_3(\mathbf{r})] &= \Phi_3(u, v, w) \\ &= \iiint \varphi_3(x, y, z) \exp\{2\pi i(ux + vy + wz)\}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z.\end{aligned} \tag{2.5.6.6}$$

The transform of projection $\varphi_2(x, y)$ along z is

$$\begin{aligned}\mathscr{F}[\varphi_2(x, y)] &= \Phi_3(u, v, 0) \\ &= \iiint \varphi_3(x, y, z) \exp\{2\pi i(ux + vy + 0z)\}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \\ &= \iiint \varphi_3(x, y, z)\, \mathrm{d}z \exp\{2\pi i(ux + vy)\}\, \mathrm{d}x\, \mathrm{d}y \\ &= \iint \varphi_2(x, y) \exp\{2\pi i(ux + vy)\}\, \mathrm{d}x\, \mathrm{d}y \\ &= \Phi_2(u, v).\end{aligned} \tag{2.5.6.7}$$

In the general case of projecting along the vector $\boldsymbol{\tau}$, the central section theorem is

$$\mathscr{F}[\varphi_2(\mathbf{x}_\tau)] = \Phi_3(\mathbf{u}_\tau), \quad \mathbf{u}_\tau \perp \boldsymbol{\tau}. \tag{2.5.6.8}$$

From this theorem it follows that the inversion of the 3D ray transform is possible if there is a continuous set of projections φ_τ corresponding to the motion of the vector $\boldsymbol{\tau}(\theta, \psi)$ over any continuous line connecting the opposite points on the unit sphere (Fig. 2.5.6.2) (Orlov's condition: Orlov, 1976). This result is evidenced by the fact that in this case the cross sections $\mathscr{F}[\varphi_2(\mathbf{x}_\tau)]$ that are perpendicular to $\boldsymbol{\tau}$ in Fourier space continuously cover the whole Fourier space, *i.e.*, they yield $\mathscr{F}[\varphi_3(\mathbf{r})]$ and thereby determine $\varphi_3(\mathbf{r}) = \mathscr{F}^{-1}[\Phi_3(\mathbf{u})]$.

In single-particle reconstruction, imaged objects are randomly and nonuniformly oriented on the substrate at different angles (Frank, 2006) and the distribution of their orientations is beyond our control; therefore, the practical impact of Orlov's condition is limited. In fact, it is more useful to determine *a posteriori*, *i.e.*, after the 3D reconstruction of the macromolecule is computed, how well the Fourier space was covered. This can be done by calculating the distribution of the 3D spectral signal-to-noise ratio (Penczek, 2002).

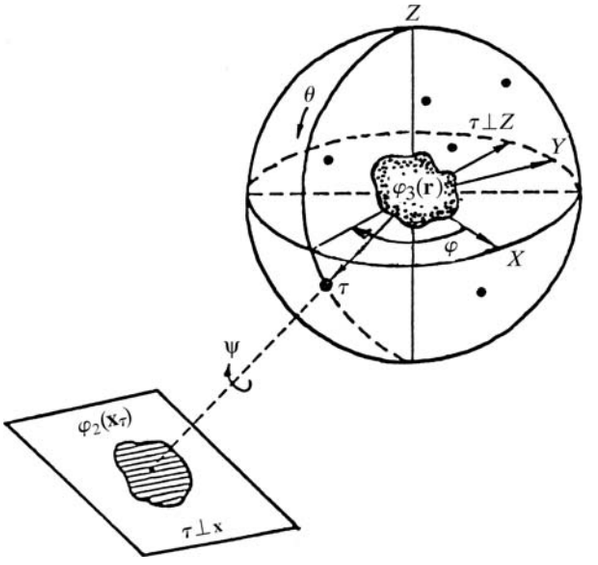


Fig. 2.5.6.2. The projection sphere and projection $\varphi_2(\mathbf{x}_\tau)$ of $\varphi_3(\mathbf{r})$ along $\boldsymbol{\tau}$ onto the plane $\boldsymbol{\tau} \perp \mathbf{x}$. The case $\boldsymbol{\tau} \perp Z$ represents orthoaxial projection. Points indicate an arbitrary distribution of projection directions $\boldsymbol{\tau}$.

2.5.6.2. 3D reconstruction in the general case

In the general case of the 3D reconstruction of $\varphi_3(\mathbf{r})$ from projections $\varphi_2(\mathbf{x}_\tau)$, the projection vector $\boldsymbol{\tau}$ occupies arbitrary positions on the projection sphere (Fig. 2.5.6.2). First, let us consider the case of a 2D function $\rho_2(\mathbf{x})$ and its ray transform $\varphi_1(x, \psi)$. We introduce an operation of backprojection b, which is stretching along $\boldsymbol{\tau}_\psi$ each 1D projection $\varphi_1(x_\psi)$ (Fig. 2.5.6.3). When the result is integrated over the full angular range of projections $\varphi_1(x, \psi)$, we obtain the projection synthesis defined as

$$b(x, y) = \int_0^\pi \varphi_1(x \cos\psi + y \sin\psi, \psi)\, \mathrm{d}\psi. \tag{2.5.6.9}$$

However, the backprojection operator is not the inverse of a 2D ray transform, as the resulting image b is blurred by the point-spread function $(x^2 + y^2)^{-1/2}$ (Vainshtein, 1971):

$$b(x, y) = \rho(x, y) * (x^2 + y^2)^{-1/2}. \tag{2.5.6.10}$$

By noting that the Fourier transform of $(x^2 + y^2)^{-1/2}$ is $(u^2 + v^2)^{-1/2}$ and by using the convolution theorem $\mathscr{F}[f * g] = \mathscr{F}[f]\mathscr{F}[g]$, we obtain the 'backprojection-filtering' inversion formula:

$$\begin{aligned}\rho(x, y) &= b(x, y) * (x^2 + y^2)^{1/2} = \mathscr{F}^{-1}\big[|\mathbf{u}|\mathscr{F}[b]\big] \\ &= \text{Filtration}_{|\mathbf{u}|}[\text{Backprojection}(\varphi_1)].\end{aligned} \tag{2.5.6.11}$$

The more commonly used 'filtered-backprojection' inversion is based on the 2D version of the central section theorem (2.5.6.8):

$$\mathscr{F}[\varphi_1(x_\psi)] = \Upsilon_2(\mathbf{u}_\psi) = \Upsilon_2(R, \Psi), \tag{2.5.6.12}$$

where $\mathscr{F}[\rho_2] = \Upsilon_2$. With this in mind, $\rho_2(\mathbf{x})$ can be related to its ray transform by evaluating the Fourier transform of ρ_2 in polar coordinates:

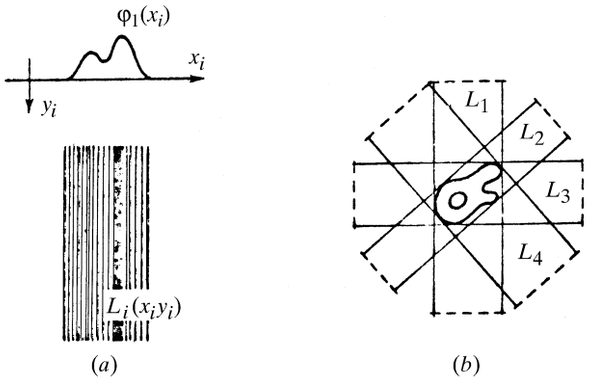


Fig. 2.5.6.3. (*a*) Formation of a backprojection function; (*b*) projection synthesis (2.5.6.9) is a superposition of these functions.

$$\begin{aligned}\rho_2(\mathbf{x}) &= \int \Upsilon_2(\mathbf{u}) \exp(-2\pi i \mathbf{u}\mathbf{x})\, \mathrm{d}\mathbf{u} \\ &= \int_0^{\pi} \int_{-\infty}^{\infty} \Upsilon_2(R, \Psi) \exp(-2\pi i \mathbf{x}\boldsymbol{\tau}) |R|\, \mathrm{d}R\, \mathrm{d}\Psi \\ &= \int_0^{\pi} \int_{-\infty}^{\infty} \mathscr{F}[\varphi_1(x_\psi)] \exp(-2\pi i \mathbf{x}\boldsymbol{\tau}) |R|\, \mathrm{d}R\, \mathrm{d}\Psi \\ &= \int_0^{\pi} \mathscr{F}^{-1}\big[|R| \mathscr{F}[\varphi_1(x_\psi)]\big]\, \mathrm{d}\Psi \\ &= \text{Backprojection}[\text{Filtration}_{|R|}(\varphi_1)]. \end{aligned} \tag{2.5.6.13}$$

In three dimensions, the backprojection stretches each 2D projection $\varphi_{2_i}[\mathbf{x}, \boldsymbol{\tau}(\theta, \psi)_i]$ along the projection direction $\boldsymbol{\tau}(\theta, \psi)_i$. A 3D synthesis is the integral over the hemisphere (Fig. 2.5.6.2)

$$b(\mathbf{r}) = \int_{\omega} \varphi_2(\mathbf{x}, \omega_{\boldsymbol{\tau}})\, \mathrm{d}\omega_{\boldsymbol{\tau}} = \varphi_3(\mathbf{r}) * (x^2 + y^2 + z^2)^{-1}. \tag{2.5.6.14}$$

Thus, in three dimensions the image b obtained using the backprojection operator is blurred by the point-spread function $1/(x^2 + y^2 + z^2)$ (Vainshtein, 1971). It is possible to derive inversion formulae analogous to (2.5.6.11) and (2.5.6.13).

The inversion formulae demonstrate that it is possible to invert the ray transform for continuous functions and for a uniform distribution of projections. In electron microscopy, the projections are never distributed uniformly in three dimensions. Indeed, a uniform distribution is not even desirable, as only certain subsets of projection directions are required for the successful inversion of a 3D ray transform, as follows from the central section theorem (2.5.6.8). In practice, we always deal with sampled data and with discrete, random and nonuniform distributions of projection directions. Therefore, the inversion formulae can be considered only as a starting point for the development of the numerical (and practical!) reconstruction algorithms. According to (2.5.6.10) and (2.5.6.14), a simple backprojection step results in reconstruction that corresponds to a convolution of the original function with a point-spread function that depends only on the distribution of projections, but not on the structure itself. Taking into account the linearity of the backprojection operation, one has to conclude that for any practically encountered distribution of projections it should be possible to derive the corresponding point-spread function and then, using either deconvolution or Fourier filtration (with a 'weighting function'), arrive at a good approximation of the structure. This observation forms the basis of the weighted backprojection algorithm (Section 2.5.6.5). Similarly, the central section theorem gives rise to direct Fourier inversion algorithms (Section 2.5.6.6). Nevertheless, since the data are discrete, the most straightforward methodology is to discretize and approach the reconstruction problem as an algebraic problem (Section 2.5.6.4).

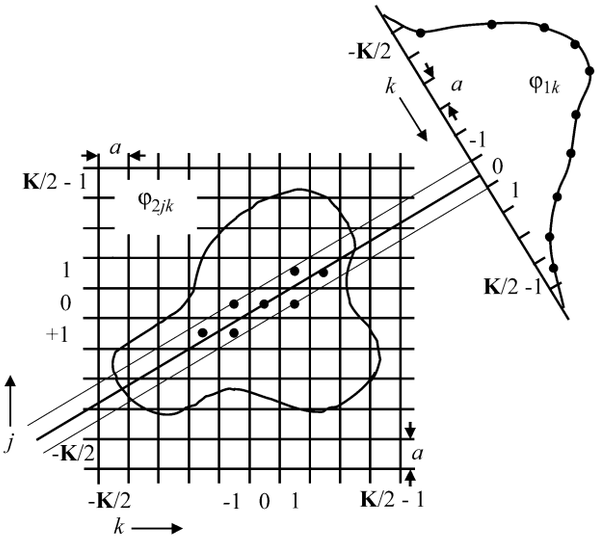


Fig. 2.5.6.4. Discretization in two dimensions ($d = 2$). The assumption is that the mass is located at the centre of the voxel.

2.5.6.3. Discretization and interpolation

In digital image processing, space is represented by a multidimensional discrete lattice. (It is sometimes expedient to use cylindrical or spherical coordinates, but these also have to be appropriately discretized.) The 2D projections $\varphi_2(\mathbf{x})$ are sampled on a Cartesian grid $\{\mathbf{k}a : \mathbf{k} \in \mathbf{Z}^d, (-\mathbf{K}/2) \leq \mathbf{k} \leq (\mathbf{K}/2)\}$, where d is the dimensionality of the grid ($d = 2$ for projections, $d = 3$ for the reconstructed object), $\mathbf{K} \in \mathbf{Z}_+^d$ is the size of the grid and a is the grid spacing (Fig. 2.5.6.4). In single-particle reconstruction, the units of a are usually ångstroms and we also assume that the data were appropriately sampled, *i.e.*, $a \leq (1/2u_{\max})$. Thus, the pixel size is less than or equal to the inverse of twice the maximum spatial frequency present in the data. Since the latter is not known in advance, a more practical rule is to select the pixel size at about one third of the expected resolution of the final structure, so that the adverse effects of interpolation are reduced.

The input electron microscopy data (projections of the macromolecule) are discretized on a 2D Cartesian grid, but each projection has a particular orientation in polar coordinates. Except for a few cases (projection directions parallel to the axes of the coordinate system of the 3D structure), an interpolation is required to relate measured samples to the voxel (volume pixel) values on the 3D Cartesian grid of the reconstructed structure (Fig. 2.5.6.4). The step of backprojection can be visualized as a set of rays with base a^{d-1} extended from projections and the ray values being added to the intersected voxels on the grid of the reconstructed structure (Figs. 2.5.6.3 and 2.5.6.4). One can select schemes that aim at approximation of the physical reality of the data collection, for example to weight the contributions by the areas of the voxels intersected by the ray or by the lengths of the lines that transverse the voxel (Huesman *et al.*, 1977). In order to reduce the time of calculations, in electron microscopy one usually assumes that all the mass is located at the centre of the voxel and the additional accuracy is achieved by application of tri- (or bi-)linear interpolation. The exception is the algebraic reconstruction technique with blobs algorithm (Marabini *et al.*, 1998), where the voxels are represented by smooth spherically symmetric volume elements [for example, the Keiser–Bessel function (2.5.6.44)].

In real space, both the projection and backprojection steps can be implemented in two different ways: as voxel driven or as ray driven (Laurette *et al.*, 2000). If we consider a projection, in the

voxel-driven approach the volume is scanned voxel by voxel. The nearest projection bin to the projection of each voxel is found, and the values in this bin and three neighbouring bins are increased by the corresponding voxel value multiplied by the weights calculated using bilinear interpolation. In the ray-driven approach, the volume is scanned along the projection rays. The value of the projection bin is increased by the values in the volume calculated in equidistant steps along the rays using trilinear interpolation. Because voxel- and ray-driven methods apply interpolation to projections or to voxels, respectively, the interpolation artifacts will be different in each case. Therefore, when calculating reconstructions using iterative algorithms that alternate between projection and backprojection steps, it is important to maintain consistency; that is, to use the same method for both steps. In either case, the computational complexity of each method is of the order of K^3, although the voxel-driven approach is faster due to the smaller number of neighbouring points used in the interpolation.

In the reconstruction methods based on the direct Fourier inversion of the 3D ray transform, the interpolation is performed in Fourier space. Unfortunately, it is difficult to design an accurate and fast interpolation scheme for the discrete Fourier space. Bilinear interpolation introduces local errors and when applied in real space it results in attenuation of high-frequency information. When applied in Fourier space, bilinear interpolation results in errors evenly spread over the whole frequency range, thus resulting in potentially severe nonlocal errors in real space. In order to eliminate this error it would be tempting to use interpolation based on Shannon's sampling theorem [Shannon, 1949; reprinted in *Proc. IEEE*, (1998), **86**, 447–457], which states that a properly sampled band-limited signal can be fully recovered from its discrete samples. For the signal represented by K^3 equispaced Fourier samples Φ_{3hkl}, the value of the Fourier transform Φ_3 at the arbitrary location (u, v, w) is given by (Crowther, Amos *et al.*, 1970)

$$\Phi_3(u, v, w) = \sum_{h=0}^{K-1} \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \Phi_{3hkl} w_h(u) w_k(v) w_l(w), \qquad (2.5.6.15)$$

where (Yuen & Fraser, 1979; Lanzavecchia & Bellon, 1994)

$$w_k(u) = \begin{cases} \dfrac{\sin\{K\pi[u-(k/K)]\}}{\sin\{\pi[u-(k/K)]\}}, & K \text{ odd}; \\ \dfrac{\sin\{K\pi[u-(k/K)]\}}{\tan\{\pi[u-(k/K)]\}}, & K \text{ even}. \end{cases} \qquad (2.5.6.16)$$

In cryo-EM, samples of Φ_3 are given at arbitrary 3D locations, as derived from Fourier transforms of 2D projection data (central section theorem) and one seeks to recover Φ_{3hkl} on the 3D Cartesian grid. Upon the inverse Fourier transform, it will yield the reconstructed object. The problem can be solved as an overdetermined system of linear equations (Crowther, DeRosier & Klug, 1970). Indeed, if we write Φ_3 and Φ_{3hkl} as 1D arrays $\mathbf{\Phi}_3$ and $\mathbf{\Phi}_{3(hkl)}$, respectively (the former has $K^2 \times$ [number of projections] elements, while the latter has K^3 elements), and we denote by $\mathbf{W}$ the appropriately dimensioned matrix of the interpolants, the least-squares solution is

$$\mathbf{\Phi}_{3(hkl)} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{\Phi}_3. \qquad (2.5.6.17)$$

The above method is impractical because of the large size of the matrix $\mathbf{W}$. In some cases, when due to symmetries the projection data are distributed approximately evenly (as in the case of icosahedral structures), the problem can be solved to a good degree of accuracy by performing the interpolation (2.5.6.17) independently along each of the three frequency axes (Crowther, DeRosier & Klug, 1970). Thus, in this case the solution to the problem of interpolation in Fourier space becomes a solution to the reconstruction problem.

For a more general single-particle reconstruction application a moving Shannon window interpolation has been proposed (Lanzavecchia & Bellon, 1994, 1998). It is based on an attenuated version of the window function and in one dimension has the form

$$\Phi_1(u) = \sum_{k=m}^{m+n-1} \Phi_{1k} \frac{\sin\{\pi[u-(k/K)]\}}{\sin\{(\pi/2)[u-(k/K)]\}} \times (\cos\{(\pi/2)[u-(k/K)]\})^A, \qquad (2.5.6.18)$$

where n is the window size ($n \ll K$) and A is an integer that is even for K odd and *vice versa*. In multidimensional cases, a product of Φ's from (2.5.6.18) is used. In the case of interpolation between equispaced samples of Φ_{2hk}, excellent results have been reported for $n = 11$ (Lanzavecchia *et al.*, 1996). However, the application to the reconstruction problem, *i.e.*, to resampling of the nonuniformly distributed Fourier data onto a 3D Cartesian grid, does not yield satisfactory results. Although general conditions under which interpolation using (2.5.6.18) can be done are known (Clark *et al.*, 1985), they are not met in practice and the results are at best nonexact. In addition, the relatively large window size required ($n = 11$) results in impractical calculation times.

2.5.6.4. The algebraic and iterative methods

The algebraic methods have been derived based on the observation that when the projection equation (2.5.6.5) is discretized, it forms a set of linear equations. Thus, pixels from all available N projections are placed (in an arbitrary order) in a vector $\varphi_{2jk}^n, n = 1, \ldots, N \rightarrow \mathbf{f}$ and the voxels of the 3D object in a vector $\varphi_{3jkl} \rightarrow \mathbf{g}$ (in an order derived from the order of $\mathbf{f}$ by algebraic relations). Note we left the exact sizes of $\mathbf{f}$ and $\mathbf{g}$ undetermined, as the major advantage of algebraic methods is that we can include in the reconstruction only pixels located within an arbitrary support in two dimensions and this support can be different for each projection; similarly, the support of the object in three dimensions can be arbitrary. Thus, the number of elements in $\mathbf{f}$ and $\mathbf{g}$ are at most K^2N and K^3, respectively, with K^2 being the number of pixels within chosen support. In the algebraic formulation the operation of projection is defined by the *projection matrix* $\mathbf{P}$ whose elements $p_{\eta\xi}$ are the interpolation weights. Their values are determined by the interpolation scheme used, but for the bi- and trilinear interpolations $0 \leq p_{\eta\xi} \leq 1$. The algebraic version of (2.5.6.5) is

$$\mathbf{f} = \mathbf{P}\mathbf{g}. \qquad (2.5.6.19)$$

Matrix $\mathbf{P}$ is rectangular and since in single-particle reconstruction the number of projections exceeds the linear size of the object ($N \gg K$) the system of equations is overdetermined. It can be solved in a least-squares sense:

$$\tilde{\mathbf{g}} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{f}, \qquad (2.5.6.20)$$

which yields a unique structure $\tilde{\mathbf{g}}$ that corresponds to the minimum of $|\mathbf{P}\mathbf{g} - \mathbf{f}|^2$. As in the case of direct inversion in Fourier space (2.5.6.17), the approach (2.5.6.20) is impractical because of the very large size of the projection matrix. Indeed, the size of $\mathbf{P}$ is $K^2N \times K^3$, which for a modest number of projections $N = 10\,000$ and image size $K^2 = 64^2 = 4096$ yields a matrix of size $\sim(4 \times 10^7) \times (2 \times 10^6)$! Nevertheless, in some cases

the projections of the structure are orthoaxial, the full 3D reconstruction reduces to a series of independent 2D reconstructions, and it becomes possible to solve (2.5.6.19) by using the singular value decomposition (SVD) of the matrix $\mathbf{P}$:

$$\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{2.5.6.21}$$

where $\mathbf{U}$ and $\mathbf{V}$ contain eigenvectors of matrices $\mathbf{P}\mathbf{P}^T$ and $\mathbf{P}^T\mathbf{P}$, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix containing the first r nonnegative square roots of the eigenvalues of $\mathbf{P}\mathbf{P}^T$ (Golub & Van Loan, 1996). The solution is given by

$$\mathbf{g} = \mathbf{V}\mathbf{\Sigma}_l^{-1}\mathbf{U}^T\mathbf{f}, \tag{2.5.6.22}$$

where $\mathbf{\Sigma}_l^{-1}$ is the inverse of the matrix containing the $l \leq r$ largest elements of $\mathbf{\Sigma}$ with the remaining values set to zero. By selecting l appropriately we achieve a measure of regularization. The advantage of the approach is that for a given geometry the decomposition has to be calculated once; thus, the method becomes very efficient if the reconstruction has to be performed repeatedly for the same distribution of projections or if additional symmetries, such as helical, are taken into account (Holmes *et al.*, 2003).

In the general 3D case, a least-squares solution can be found using one of the iterative approaches that take advantage of the fact that the projection matrix is sparse. Indeed, if bilinear interpolation is used, a row of $\mathbf{P}$ will contain only about $4N$ nonzero elements (out of the total $\sim K^2N$). The main idea is that the matrix $\mathbf{P}$ is not explicitly calculated or stored; instead, its elements are calculated repeatedly during iterations as needed. In the simultaneous iterative reconstruction technique (SIRT) we find the minimum of

$$L(\mathbf{g}) = |\mathbf{P}\mathbf{g} - \mathbf{f}|^2 \tag{2.5.6.23}$$

by selecting the initial 3D structure $\mathbf{g}^0$ (usually set to zero) and by iteratively updating its current approximation $\mathbf{g}^{i+1}$ using the gradient of the objective function (2.5.6.23) $\nabla L(\mathbf{g})$,

$$\mathbf{g}^{i+1} = \mathbf{g}^i - \lambda^i\mathbf{P}^T(\mathbf{P}\mathbf{g}^i - \mathbf{f}) = \mathbf{g}^i - \lambda^i(\mathbf{P}^T\mathbf{P}\mathbf{g}^i - \mathbf{P}^T\mathbf{f}). \tag{2.5.6.24}$$

Setting the relaxation parameter $\lambda^i = \lambda = \text{constant}$ yields Richardson's algorithm (Gilbert, 1972). In single-particle reconstruction applications convergence (*i.e.*, a solution that is not dominated by the noise) is usually reached in ~100 iterations.

SIRT is extensively used in single-particle reconstruction (Frank *et al.*, 1996) because it yields superior results over a wide range of experimental conditions (Penczek *et al.*, 1992) and, when presented with angular gaps in the distribution of projections, produces the least disturbing artifacts. Note that strictly speaking, filtered backprojection and direct Fourier inversion algorithms are not applicable to data that do not cover the Fourier space fully. In addition, SIRT offers considerable flexibility in EM applications. First, it is possible to accelerate the convergence by adjusting relaxation parameters: setting $\lambda^i = \arg\min_{\lambda \geq 0} L(g^i - \lambda\nabla L(g^i))$ results in a steepest decent algorithm. Secondly, even faster convergence (in ~10 iterations) is achieved by solving (2.5.6.23) using the conjugate-gradient method, but this requires addition of a regularizing term to (2.5.6.23) in order to prevent excessive enhancement of noise. Such a term has the form $|\mathbf{B}\mathbf{g}|^2$, where matrix $\mathbf{B}$ is a discrete approximation of a Laplacian or higher-order derivatives. Thirdly, it is possible to take into account the underfocus settings of the microscope by including the contrast transfer function (CTF) of the microscope in (2.5.6.23) and by solving the problem for the structure $\mathbf{g}$ that in effect will be corrected for the CTF,

$$L(\mathbf{g}) = (1-\eta)\sum_\kappa |\mathbf{S}_\kappa\mathbf{P}_\kappa\mathbf{g} - \mathbf{f}_\kappa|^2 + \eta|\mathbf{B}\mathbf{g}|^2. \tag{2.5.6.25}$$

Here we assumed that the projection data were grouped according to the defocus settings indexed by κ and we introduced a Lagrange multiplier η whose value determines the smoothness of the solution $\mathbf{g}$ (Zhu *et al.*, 1997). $\mathbf{S}_\kappa$ represents the space-invariant point-spread function of the microscope for the κth defocus group; thus $\mathbf{S}_\kappa$ has a block-Toeplitz structure (Biemond *et al.*, 1990) and $\mathbf{S}_\kappa = \mathbf{S}_\kappa^T$. If we assume that there is no astigmatism (which means that the point-spread function is rotationally invariant), $\mathbf{S}_\kappa$ can be either applied to the projection of the structure or to the 3D structure itself. With these assumptions, (2.5.6.25) can be solved using the iterative scheme (Penczek *et al.*, 1997)

$$\mathbf{g}^{i+1} = \mathbf{g}^i - \lambda\big\{(1-\eta)\big[\sum_\kappa(\mathbf{S}_\kappa^T\mathbf{P}_\kappa^T\mathbf{P}_\kappa\mathbf{S}_\kappa\mathbf{g}^i) - \sum_\kappa(\mathbf{S}_\kappa^T\mathbf{P}_\kappa^T\mathbf{f}_\kappa)\big] + \eta\mathbf{B}^T\mathbf{B}\mathbf{g}^i\big\}. \tag{2.5.6.26}$$

The second sum can be precalculated and stored as a 3D volume in the computer memory. Thus, the input projections have to be read only once and are never accessed again during the course of the iterations, which eliminates the need to store them in the memory [as is also the case in (2.5.6.24)]. In addition, the product $\mathbf{B}^T\mathbf{B}$ is the 3D Laplacian, which can be applied efficiently without actually creating the matrix $\mathbf{B}^T\mathbf{B}$. Finally, because of the large size of the matrix $\mathbf{S}_\kappa$ it is more convenient to apply it in Fourier space and to modify the Fourier transform of the volume instead of using matrix multiplication or real-space convolution. Algorithms (2.5.6.24)–(2.5.6.26) are implemented in the *SPIDER* package (Frank *et al.*, 1996) and (2.5.6.24) in the *SPARX* package (Hohn *et al.*, 2007).

The algebraic reconstruction technique (ART) predates SIRT; in the context of tomographic reconstructions it was proposed by Gordon and co-workers (Gordon *et al.*, 1970) and later it was recognized as a version of Kaczmarz's method for iteratively solving (2.5.6.23) (Kaczmarz, 1993). We write (2.5.6.19) as a set of systems of equations, each relating single pixels f_n, $n = 1, \ldots, NK^2$ in projections with the 3D structure,

$$f_n = \mathbf{p}_n^T\mathbf{g}, \quad n = 1, \ldots, NK^2. \tag{2.5.6.27}$$

Note (2.5.6.27) and (2.5.6.19) are equivalent, as $\mathbf{f} = [f_1 f_2 \ldots f_N]^T$ and $\mathbf{P} = [\mathbf{p}_1\mathbf{p}_2 \ldots \mathbf{p}_N]^T$. With this notation and a relaxation parameter $0 < \mu < 2$, ART comprises the following steps:

(1) set $i = 0$ and the initial structure $\mathbf{g}^0$;

(2) for $n = 1, \ldots, NK^2$, set

$$\mathbf{g}^{iNK^2+n} = \mathbf{g}^{iNK^2+n-1} - \mu(\mathbf{p}_n^T\mathbf{g}^{iNK^2+n-1} - f_n)(\mathbf{p}_n/|\mathbf{p}_n|); \tag{2.5.6.28}$$

(3) set $i = i + 1$; go to step (2).

Although the mathematical forms of update equations in SIRT (2.5.6.24) and in ART (2.5.6.28) are very similar, there are profound differences between them. In SIRT, all voxels in the structure are corrected simultaneously after projections and backprojections of the current update of the structure are calculated. In ART, the projection/backprojection step (2.5.6.28) involves only correction with respect to an individual pixel in a single projection immediately followed by the update of the structure. This results in a much faster convergence of ART as compared to SIRT. Further acceleration can be achieved by

selecting the order in which pixels enter the correction step (2) in (2.5.6.28). It was observed that if a pixel is selected such that its projection direction is perpendicular to the projection direction of the previous pixel, the convergence is achieved faster (Hamaker & Solmon, 1978; Herman & Meyer, 1993). Interestingly, a random order works almost equally well (Natterer & Wübbeling, 2001).

In single-particle reconstruction, ART has been introduced in the form of 'ART with blobs' (Marabini *et al.*, 1998) and is available in the *Xmipp* package (Sorzano *et al.*, 2004). In this implementation, the reconstruction structure is represented by a linear combination of spherically symmetric, smooth, spatially limited basis functions, such as Kaiser–Bessel window functions (Lewitt, 1990, 1992; Matej & Lewitt, 1996). Introduction of blobs significantly reduces the number of iterations necessary to reach an acceptable solution (Marabini *et al.*, 1998).

The major advantage of iterative reconstruction methods is the ability to take advantage of *a priori* knowledge, *i.e.*, any information about the protein structure that was not initially included in the data processing, and introduce it into the reconstruction process in the form of constraints. Examples of such constraints include similarity to the experimental (measured) data, positivity of the protein mass density (only valid in conjunction with the CTF correction), bounded spatial support *etc*. Formally, the process of enforcing selected constraints is best described in the framework of the projections onto convex sets (POCS) theory (Sezan & Stark, 1982; Youla & Webb, 1982; Sezan, 1992) introduced into EM by Carazo and co-workers (Carazo & Carrascosa, 1986, 1987; Carazo, 1992).

2.5.6.5. Filtered backprojection

The method of filtered backprojection (FBP) is based on inversion formulae (2.5.6.11) (in two dimensions) or (2.5.6.14) (in three dimensions). It comprises the following steps: (i) a Fourier transform of each projection is computed; (ii) Fourier transforms of projections are multiplied by filters that account for a particular distribution of projections in Fourier space; (iii) the filtered projections are inversely Fourier transformed; (iv) real-space backprojection of processed projections yields the reconstruction. The method is particularly attractive due to the fact that the reconstruction calculated using simple real-space backprojection can be made efficient if the filter function is easy to compute.

In two dimensions with uniformly distributed projections the weighting function $c(|R|, \Psi)$ in Fourier space is the 'ramp function' $|R|$ [(2.5.6.13)]. In two dimensions with nonuniformly distributed projections, when the analytical form of the distribution of projections is not known, an appropriate approximation to $c(|R|, \Psi)$ has to be found. A good choice is to select weights such that the backprojection integral becomes approximated by a Riemann sum (Penczek *et al.*, 1996),

$$c(|R|, \Psi) = |R|\,\mathrm{d}R\,\mathrm{d}\Psi \rightarrow c(R_j, \Psi_i) = R_j \frac{1}{2\pi}\frac{\Psi_{i+1} - \Psi_{i-1}}{2} = R_j \frac{\Delta\Psi_i}{4\pi}. \quad (2.5.6.29)$$

For a given set of angles the weights $c(R_j, \Psi_i)$ are easily computed (Fig. 2.5.6.5).

In three dimensions, the weighting (2.5.6.29) is applicable in a single-axis tilt data-collection geometry, where the 3D reconstruction can be calculated as a series of independent 2D reconstructions. In the general 3D case, the analogue of weighting (2.5.6.29) cannot be used, as the data are given in the form of 2D projections and it is not immediately apparent what fraction of the 3D Fourier volume is occupied by Fourier pixels in projections. However, the analogue of weighting (2.5.6.29) can be used in the inversion of 3D Radon transforms or in a direction

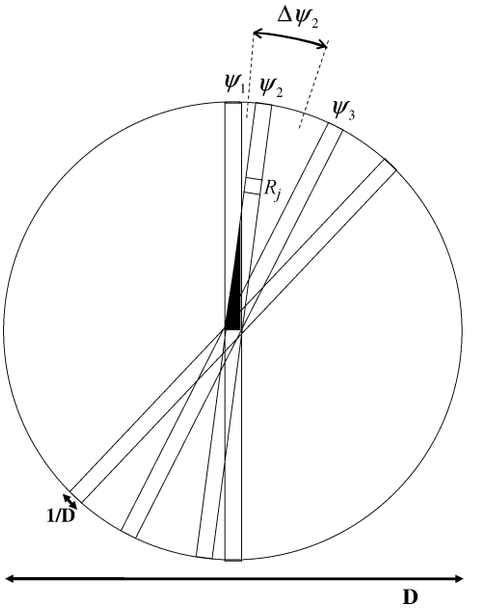


Fig. 2.5.6.5. Nonuniform distribution of projections. The projection weights for the reconstruction algorithms are chosen such that the backprojection integral becomes approximated by a Riemann sum and are equal to the angular length of an arc $\Delta\Psi_i$ (2.5.6.29). In Fourier space, projections of an object with real-space radius D form rectangles with width $1/D$. In the exact filter backprojection reconstruction method, the weights are derived based on the amount the overlap between projections (2.5.6.29).

inversion of 3D ray transforms that is based on an intermediate step of converting 2D projection data to 1D projection data, as described in Section 2.5.6.6.

In order to arrive at a workable solution, the weighting functions applicable to 2D projections are constructed based on an explicitly or implicitly formulated concept of the 'local density' of projections. This concept was introduced by Bracewell (Bracewell & Riddle, 1967), who suggested for a 2D case of nonuniformly distributed projections a heuristic weighting function,

$$c(R_j, \Psi_i) = \frac{R_j}{\sum_l \exp\left[-\text{constant}(|\Psi_i - \Psi_l| \bmod \pi)^2\right]}. \quad (2.5.6.30)$$

The weighting function (2.5.6.30) can be easily extended to three dimensions; however, it has a major disadvantage that for a uniform distribution of projections it does not approximate well the weighting function (2.5.6.29), which we consider optimal.

Radermacher *et al.* (1986) proposed a derivation of a general weighting function using a deconvolution kernel calculated for a given (nonuniform) distribution of projections and, in modification of (2.5.6.14), a finite length of backprojection (Fig. 2.5.6.3). Such a 'truncated' backprojection is

$$\hat{b}_i(\mathbf{r}) = \varphi_{2_i}(\mathbf{x}_\tau) * l(\mathbf{r}), \quad \boldsymbol{\tau} \perp \mathbf{x} \quad (2.5.6.31)$$

with projection directions $\boldsymbol{\tau}(\theta, \psi)_i$ and

$$l(\mathbf{r}) = \delta(\mathbf{x}_\tau) t(z_\tau), \qquad t(z_\tau) = \begin{cases} 1 & -(D/2) \le z \le (D/2), \\ 0 & \text{otherwise}, \end{cases} \quad (2.5.6.32)$$

where D is the diameter of the object or the length of the backprojection l. By taking the Fourier transform of (2.5.6.31) and using the central section theorem (2.5.6.8), we obtain a 3D Fourier transform of the backprojected projection,

$$\Phi_{3_i}(u, v, w) = \Phi_{2_i}(\mathbf{u}_\tau) D \operatorname{sinc}(D\pi w_\tau), \tag{2.5.6.33}$$

and the 3D reconstruction is obtained by the inverse Fourier transform of the sum of contributions given by (2.5.6.33),

$$\mathscr{F}[\hat{b}(\mathbf{r})] = \sum_i \Phi_{2_i}(\mathbf{u}_\tau) D \operatorname{sinc}(D\pi w_\tau). \tag{2.5.6.34}$$

z_τ and w_τ are variables in real and Fourier spaces, respectively, both extending in the direction of the projection direction $\boldsymbol{\tau}(\theta, \psi)_i$.

In this analysis, the transfer function of the backprojection algorithm is obtained by setting $\Phi_{2_i}(\mathbf{u}_\tau) = 1$, that is by finding the response of the algorithm to the input composed of delta functions in real space. This yields the inversion formula for a general weighted backprojection algorithm:

$$\varphi_3(\mathbf{r}) = \mathscr{F}[\hat{b}(\mathbf{r})] c(u, v, w), \tag{2.5.6.35}$$

with the weighting function given by

$$c(u, v, w) = \frac{1}{\sum_i D \operatorname{sinc}(D\pi w_\tau)}. \tag{2.5.6.36}$$

The general weighting function (2.5.6.36) is consistent with analytical solutions (2.5.6.11) and (2.5.6.14), as it can be shown that by assuming infinite support ($D \rightarrow \infty$) and continuous and uniform distribution of projection directions, in two dimensions one obtains $c(u, v) = (u^2 + v^2)^{-1/2}$ (Radermacher, 2000).

The derivation of (2.5.6.36) is based on the analysis of continuous functions and its direct application to discrete data results in reconstruction artifacts; therefore, Radermacher (1992) proposed attenuating the sinc functions in (2.5.6.36) by exponent functions with decay depending on the diameter of the structure D, or simply replacing the sinc functions by exponent functions. This, however, reduces the concept of the weighting function corresponding to the deconvolution to the concept of the weighting function representing the 'local density' of projections (2.5.6.30). The general weighted backprojection reconstruction is implemented in *SPIDER* (using exponent-based weighting functions) (Frank *et al.*, 1996), *Xmipp* (Sorzano *et al.*, 2004) and *SPARX* (Hohn *et al.*, 2007).

Harauz & van Heel (1986) proposed basing the calculation of the density of projections, thus the weighting function, on the overlap of Fourier transforms of projections in Fourier space. Although the concept is general, it can be easily approached in two dimensions. If the diameter of the object is D, the width of a Fourier transform of a projection is $2/D$ (Fig. 2.5.6.5), as follows from the central section theorem (2.5.6.8). Harauz and van Heel postulated that the weighting should be inversely proportional to the sum of geometrical overlaps between a given central section and the remaining central sections. For a pair of projections il, this overlap is

$$o_{il}(R) = T[DR \sin(\Psi_i - \Psi_l)], \tag{2.5.6.37}$$

where T represents the triangle function (selected by the authors because it can be calculated efficiently). Also, owing to the Friedel symmetry of central sections, the angles in (2.5.6.37) are restricted such that $0 \leq \Psi_i - \Psi_l \leq (\pi/2)$. In this formulation, the overlap is limited to the maximum frequency,

$$R_{il}^{\max} = \frac{2}{D \sin(\Psi_i - \Psi_l)}. \tag{2.5.6.38}$$

Thus, the overlap function becomes

$$o_{il}(R) = \begin{cases} 1 - (R/R_{il}^{\max}) & 0 \leq R < u_{il}^{\max} \\ 0 & R > u_{il}^{\max} \end{cases}. \tag{2.5.6.39}$$

In effect, the weighting function, called by the authors an 'exact filter', is

$$c(R, \Psi_i) = \frac{1}{1 + \sum_{l, l \neq i} o_{il}(R)}. \tag{2.5.6.40}$$

The weighting function (2.5.6.40) easily extends to three dimensions; however, the calculation of the overlap between central sections in three dimensions (represented by slabs) is more elaborate (Harauz & van Heel, 1986). The method is conceptually simple and computationally efficient. However, (2.5.6.40) does not approximate the correct weighting well for a uniform distribution of projections [*i.e.*, it should yield $c(R, \Psi_i) = R$]. This, as can be seen by integrating (2.5.6.39) over the whole angular range, is not the case. The exact filter backprojection reconstruction is implemented in the *IMAGIC* (van Heel *et al.*, 1996), *SPIDER* (Frank *et al.*, 1996) and *SPARX* (Hohn *et al.*, 2007) packages.

The 3D reconstruction methods based on filtered backprojection are commonly used in single-particle reconstruction. The reasons are: their versatility, ease of implementation, and – in comparison with iterative methods – good computational efficiency. Unlike in iterative methods, there are no parameters to adjust, although it has been noted that the results depend on the value of the diameter D of the structure in all three weighting functions [(2.5.6.30), (2.5.6.36) and (2.5.6.38)], so the performance of the reconstruction algorithm can be optimized for a particular data-collection geometry by changing the value of D (Paul *et al.*, 2004). However, because computation of the weighting function involves calculation of pairwise distances between projections, the computational complexity is proportional to the square of the number of projections and for large data sets these methods become inefficient. It also has to be noted that the weighting functions (2.5.6.30), (2.5.6.36) and (2.5.6.40) remain approximations of the correct weighting function (2.5.6.29).

2.5.6.6. Direct Fourier inversion

Direct Fourier methods are based on the central section theorem (2.5.6.8). A set of the 2D Fourier transforms of projections yields an approximation to Φ_3 on a nonuniform 3D grid, and a subsequent numerical 3D inverse Fourier transform gives an approximation to φ_3. If the 3D inverse Fourier transform could be realized by means of the 3D inverse fast Fourier transform (FFT), one would have a very fast reconstruction algorithm. Unfortunately, the preprocessing step yields Φ_3 on a nonuniform grid. In effect, the 3D inverse FFT is not applicable and an additional step of recovering samples of Φ_3 on a uniform grid from the available samples on a nonuniform grid is necessary.

One possibility is to resample the nonuniformly sampled version of Φ_3 onto a 3D Cartesian grid by some form of interpolation. For example, Grigorieff used a modified trilinear interpolation scheme in the *FREALIGN* package (Grigorieff, 1998). Simple interpolation methods have been found to give inaccurate results, although more sophisticated interpolation schemes can go a long way to improve the accuracy (Lanza-

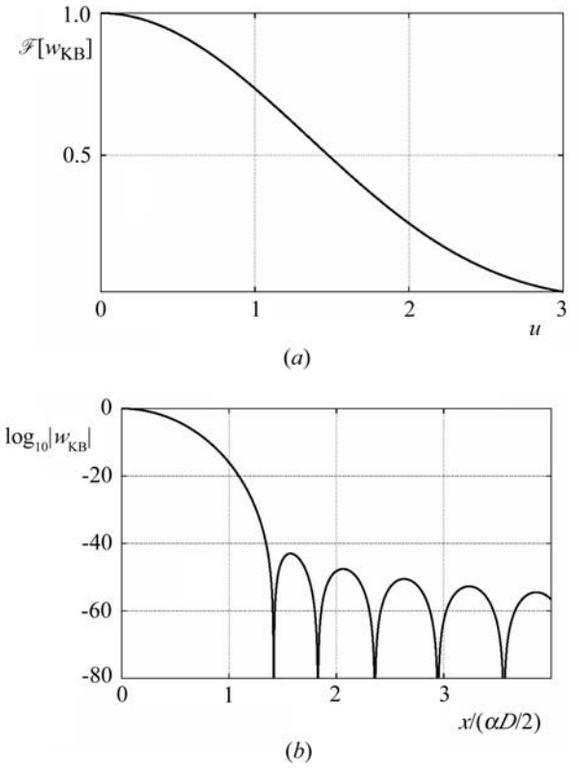


Fig. 2.5.6.6. Kaiser–Bessel window function used in the gridding reconstruction algorithm GDFR. (*a*) In Fourier space, the window function is effectively zero outside the support of six Fourier pixels. (*b*) In real space, the zeros of the window function are beyond the radius of the reconstructed object.

vecchia *et al.*, 1993). Unfortunately, the recommended window size makes them impractical for most applications.

The most accurate Fourier reconstruction methods are those that employ nonuniform Fourier transforms, particularly the 3D gridding method (O'Sullivan, 1985; Schomberg & Timmer, 1995). The gridding-based direct Fourier reconstruction algorithm (GDFR) (Penczek *et al.*, 2004) was developed specifically for single-particle reconstruction. It comprises three steps:

(1) The first step, known as 'gridding', involves calculating for the Fourier transform of each projection the convolution

$$\sum_i \mathscr{F}[w] * (c\mathscr{F}[\varphi_{2_i}]), \tag{2.5.6.41}$$

where c are 'gridding weights' designed to compensate for the nonuniform distribution of the grid points and $\mathscr{F}[w]$ is an appropriately chosen convolution kernel. After processing all projections, this step yields samples of $\mathscr{F}[w] * \mathscr{F}[\varphi_3]$ on a Cartesian grid.

(2) 3D inverse FFT is used to compute

$$w\varphi_3 = \mathscr{F}^{-1}\big[\mathscr{F}[w] * \mathscr{F}[\varphi_3]\big] \tag{2.5.6.42}$$

on a Cartesian grid.

(3) The weights are removed using division:

$$\varphi_3 = w\varphi_3/w. \tag{2.5.6.43}$$

The method involves a number of parameters. First, we need to decide the oversampling factor for the padding with zeros of input projections before FFTs are computed. In GDFR, the oversampling is set to two, although smaller factors can also yield good results. Second, we need a *window function* $\mathscr{F}[w]$ whose support in Fourier space is 'small'. In order to assure good computational efficiency of the convolution step (2.5.6.41), in GDFR this support was set to six Fourier voxels. In addition, in order to prevent division by zeros in (2.5.6.43), the *weighting function* $w = \mathscr{F}^{-1}[\mathscr{F}[w]]$ must be positive within the support of the reconstructed object. A recommendable window function is the separable, bell-shaped Kaiser–Bessel window (O'Sullivan, 1985; Jackson *et al.*, 1991; Schomberg & Timmer, 1995) (Fig. 2.5.6.6*a*):

$$\mathscr{F}[w_{KB}(\mathbf{u})] = \begin{cases} \displaystyle\prod_{v=1}^{d} \frac{I_0\{\pi D\alpha s_v[1-(u_v/s_v)^2]^{1/2}\}}{2s_v}, & \mathbf{u} \in [-\mathbf{s}, \mathbf{s}], \\ 0, & \mathbf{u} \notin [-\mathbf{s}, \mathbf{s}], \end{cases} \tag{2.5.6.44}$$

where I_0 is the zero-order modified Bessel function and $\alpha = 1.25$ is a parameter. The weighting function associated with $\mathscr{F}[w_{KB}]$ is

$$w_{KB}(\mathbf{x}) = \prod_{v=1}^{d} \frac{\sinh\Big[\pi D\alpha s_v\big(1-\{x_v/[\alpha(D/2)]^2\}\big)^{1/2}\Big]}{\pi D\alpha s_v\big(1-\{x_v/[\alpha(D/2)]^2\}\big)^{1/2}}. \tag{2.5.6.45}$$

Finally, the gridding weights c are chosen such that in discrete implementation of (2.5.6.41) and (2.5.6.42) we obtain a Riemann sum approximating the respective integral (Schomberg, 2002). The Voronoi diagram (Okabe *et al.*, 2000) of the sampling points provides a good partitioning of sampling space, so the gridding weights c are computed by constructing a Voronoi diagram of the grid points and by choosing the weights as the volumes of the Voronoi cells (Fig. 2.5.6.7*a*). In cryo-EM, the number of projections, thus the number of sampling points in Fourier space, is extremely large. Thus, although the calculation of the gridding weights *via* the 3D Voronoi diagram of the nonuniformly spaced grid points for $\mathscr{F}[\varphi_3]$ would lead to an accurate direct Fourier method, the method would be very slow and would require excessive computer memory. To circumvent this problem, in GDFR (Penczek *et al.*, 2004) the 2D reverse gridding method is used to compute the Fourier transform of each projection on a 2D polar grid. In this way, $\mathscr{F}[\varphi_3]$ is obtained on a 3D spherical grid, where the grid points are located both on centred spheres and on straight lines through the origin. Accordingly, it becomes possible to partition the sampled region into suitable sampling cells *via* the computation of a 2D Voronoi diagram on a unit sphere, rather than a 3D Voronoi diagram in Euclidean space (Fig. 2.5.6.7*b*). This significantly reduces the memory requirements and improves the computational efficiency of the method, particularly when a fast $O(n \log n)$ algorithm for computing the Voronoi diagram on the unit sphere is employed (Renka, 1997).

The gridding weights in GDFR correspond to the weighting functions in filtered backprojection algorithms. Moreover, setting the gridding weights to be equal to angular regions on the unit sphere directly corresponds to the equating of weighting function (2.5.6.29) in 2D filtered backprojection to the length of an arc on a unitary circle. Thus, both methods yield nearly optimum projection density functions.

The reversed gridding method is obtained by reversing the sequence of steps (1)–(3) of the gridding method:

(1) the input image is divided by the weighting function φ_2/w;

(2) the image is padded with zeros and 2D FFT is used to compute $\mathscr{F}[\varphi_2/w]$;

(3) gridding is used to compute $\mathscr{F}[w] * \mathscr{F}[\varphi_2/w]$ on an arbitrary nonuniform grid.

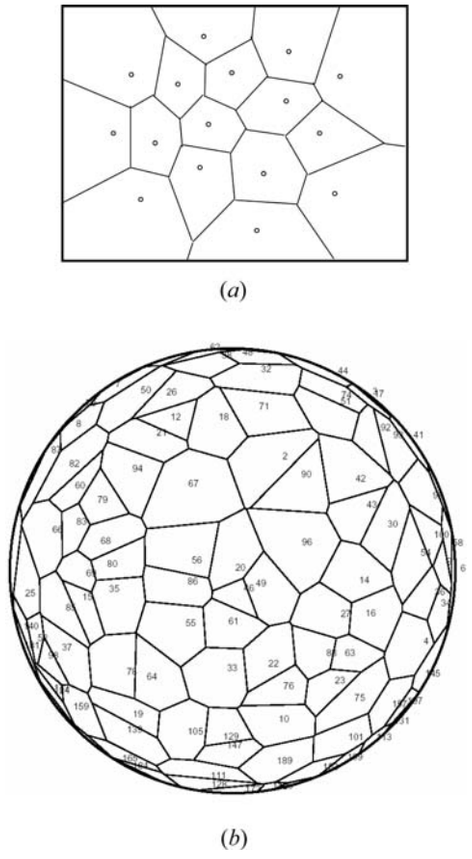


Fig. 2.5.6.7. Partitions of the sampling Fourier space using Voronoi diagrams in direct inversion algorithms. (*a*) Central section of a 3D Fourier volume with sampling points originating from 2D Fourier transforms of projections. Although 2D projections are sampled on a uniform (Cartesian) grid, the arbitrary rotations of projections in 3D space yields a nonuniform distribution of points in three dimensions. In effect, the 3D reconstruction by direct inversion using 3D FFT is not possible. (*b*) Voronoi diagram on a sphere in the GDFR algorithm. Using the reverse gridding method, the 2D Fourier transforms of projections are resampled onto 1D central lines using a constant angular step. In 3D Fourier space, they are located on central sections and their angular directions are evenly distributed on grand circles. However, since central sections have nonuniform distributions, the distribution of angular directions (sampling points on the unitary sphere) is also nonuniform and effectively random.

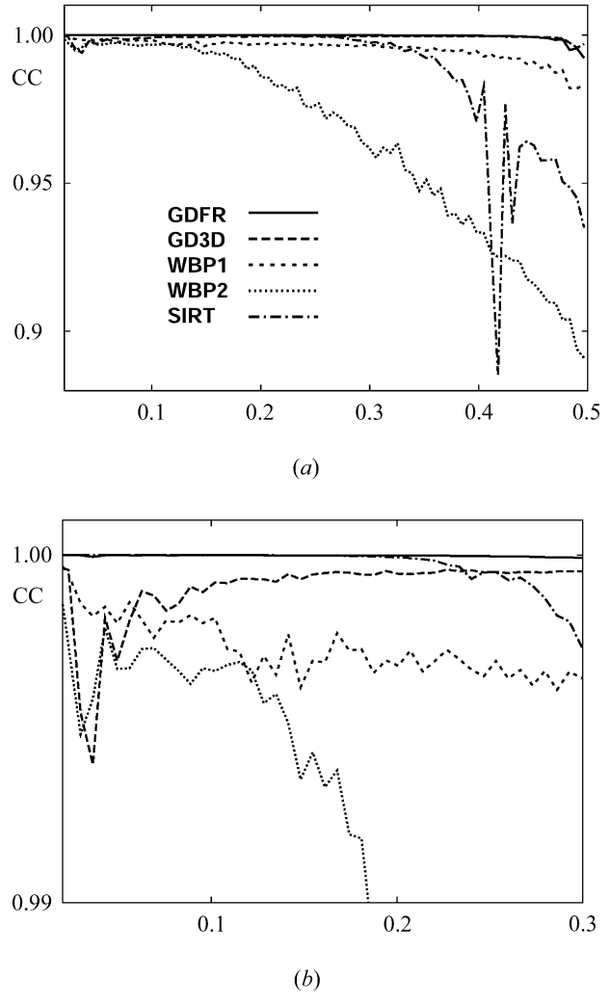


Fig. 2.5.6.8. (*a*) Plot of correlation coefficients (CCs) calculated between Fourier transforms of the reconstructed structure and the original phantom as a function of the magnitude of the spatial frequency using five reconstruction algorithms. GDFR: gridding direct Fourier reconstruction with Voronoi weights; GD3D: gridding reconstruction with simplified gridding weights; WBP1: general weighted backprojection with exponent-based weighting function; WBP2: exact filter weighted backprojection; SIRT: simultaneous iterative reconstruction algorithm. Noise-free projection data were computed in Fourier space using the reverse gridding method. (*b*) Rescaled version of the low-frequency range of (*a*). Note the different scales in (*a*) and (*b*). The horizontal axis is scaled in absolute frequency units with 0.5 equal to the Nyquist frequency.

Note the reverse gridding does not require explicit gridding weights, as in the third step they are constant (Penczek *et al.*, 2004).

In GDFR, the third step of the reversed gridding results in a set of 1D Fourier central lines $\mathcal{F}[\varphi_{1l}]$ calculated using a constant angular step. Clearly, upon inverse Fourier transform they amount to a Radon transform $\varphi_1(u, \psi)$ of the projection and, upon repeating the process for all available projections, they yield a Radon transform $\varphi_1(u, \boldsymbol{\theta})$ of a 3D object φ_3 (albeit non-uniformly sampled with respect to Eulerian angles). Thus GDFR, in addition to being a method of inverting a 3D ray transform, is also a highly accurate method of inverting a 3D Radon transform. GDFR is implemented in the *SPIDER* package (Frank *et al.*, 1996).

The results of a comparison of selected reconstruction algorithms are shown in Fig. 2.5.6.8. The tests were performed using simulated data with the projections of a phantom 3D structure calculated using the inverse gridding method (Penczek *et al.*, 2004). The GDFR yields a virtually perfect reconstruction that agrees with the phantom over the whole frequency range (with the exception of the highest frequencies, but these cannot be reproduced due to geometrical limitations). GD3D is also a gridding-based reconstruction algorithm, in which the gridding weights are calculated directly from contributions of the weighting function to the 3D Fourier pixels and are applied to the voxels in the 3D Fourier volume (Jackson *et al.*, 1991). This yields an approximation to the 'local density' of contributing projections. The simplified gridding weights in GD3D result in deterioration of the reconstruction in the low- and intermediate-frequency ranges. The general weighted backprojection with exponent-based weighting function (WBP1) performs quite well, although reproduction in the low-frequency range is inferior. Similar artifacts are present in the reconstruction using the exact filter weighted backprojection (WBP2), which, in addition, performs disappointingly at higher frequencies. This relatively poor performance is attributed to the nonoptimal weighting schemes used in both methods. It is also interesting to note that the backprojection step is identical in both algorithms; they only differ by the weighting function. The significant difference in their performance attests to the importance of good weighting schemes for high-quality 3D reconstructions from nonuniformly distributed projections. SIRT yields a reconstruction that in the low- and intermediate-frequency ranges matches in quality the reconstruction obtained with GDFR. However, there is a significant loss of quality at high frequencies. This is due to the

inferior linear interpolation scheme used in SIRT (this impediment is shared with WBP1 and WBP2). [The dip of the SIRT fidelity curve at a spatial frequency of 0.42 in Fig. 2.5.6.8(*a*) is due to an inconsistency between the interpolation methods used internally in SIRT to generate intermediate projections of the object to be reconstructed and the Fourier-space-based method used to generate the test data.] In terms of computational efficiency, gridding-based algorithms outperform weighted back-projection algorithms by a small factor while SIRT is approximately ten times slower (depending on the number of iterations used).

2.5.6.7. 3D reconstruction of symmetric objects

Many objects imaged by EM have symmetries; the two types that are often met are helical (phage tails, *F*-actin, microtubules, myosin thick filaments, and bacterial pili and flagella) and point-group symmetries (many macromolecular assemblies, virus capsids). If the object has helical symmetry, it is convenient to use cylindrical coordinates and a dedicated reconstruction algorithm; particularly for the initial analysis of the data that has to be done in Fourier space. Diffraction from such structures with c periodicity and scattering density $\varphi(r, \psi, z)$ is defined by the Fourier–Bessel transform:

$$\begin{aligned}\Phi(R, \Psi, Z) &= \sum_{-\infty}^{+\infty} \exp\Big[in\Big(\Psi + \frac{\pi}{2}\Big)\Big] \int_0^\infty \int_0^{2\pi} \int_0^l \varphi(r, \psi, z) \\ &\quad \times J_n(2\pi r R) \exp[-i(n\psi + 2\pi z Z)] r \, \mathrm{d}r \, \mathrm{d}\psi \, \mathrm{d}z \\ &= \sum_n G_n(R, Z) \exp\Big[in\Big(\Psi + \frac{\pi}{2}\Big)\Big]. \qquad (2.5.6.46)\end{aligned}$$

The inverse transform has the form

$$\rho(r, \psi, z) = \sum_n \int g_n(r, Z) \exp(in\psi) \exp(2\pi i z Z) \, \mathrm{d}Z, \qquad (2.5.6.47)$$

so that g_n and G_n are the mutual Bessel transforms

$$G_n(R, Z) = \int_0^\infty g_n(r, Z) J_n(2\pi r R) 2\pi r \, \mathrm{d}r \qquad (2.5.6.48)$$

and

$$g_n(r, Z) = \int_0^\infty G_n(R, Z) J_n(2\pi r R) 2\pi R \, \mathrm{d}R. \qquad (2.5.6.49)$$

Owing to helical symmetry, (2.5.6.48) and (2.5.6.49) contain only those of the Bessel functions that satisfy the selection rule (Cochran *et al.*, 1952)

$$l = mp + (nq/N), \qquad (2.5.6.50)$$

where N, q and p are the helix symmetry parameters, $m = 0, \pm 1, \pm 2, \ldots$. Each layer l is practically determined by the single function J_n with the lowest n; the contributions of other functions are neglected. Thus, the Fourier transform of one projection of a helical structure, with an account of symmetry and phases, gives the 3D transform (2.5.6.49). However, biological specimens tend to be flexible and disordered, and exact helical symmetry is rarely observed. A possible approach to dealing with flexibility is to computationally straighten filaments (Egelman, 1986), but this has the potential for introducing artifacts. Another difficulty with helical analysis is that the indexing of a pattern can be ambiguous and the wrong symmetry can be chosen (Egelman & Stasiak, 1988). Further complications exist when the filament does not have a precisely defined helical symmetry, such as *F*-actin, which has a random variable twist (Egelman *et al.*, 1982). To address these problems, Egelman developed a real-space refinement method for the reconstruction of helical filaments that is capable of determining the helical symmetry of an unknown structure (Egelman, 2000). In this approach, the Fourier–Bessel reconstruction algorithm is replaced by a general reconstruction algorithm discussed in Section 2.5.6.6 with the real-space projections supplied as input. Thus, the symmetry is not enforced within the reconstruction algorithm; instead, it is determined and imposed subsequently by real-space averaging.

The presence of point-group symmetry in the structure means that any projection yields as many symmetry-related (but differently oriented) copies of itself as the number of symmetry operations in the group. For example, for three dimensions, each projection enters the reconstruction process at six different projection directions, while for icosahedral symmetry (I) the number is sixty! This high multiplicity makes it possible to write a reconstruction program that explicitly takes into account the symmetry and can perform the task much faster than generic algorithms. Such programs are usually part of dedicated software packages that were specifically designed for the determination of icosahedral structures (Fuller *et al.*, 1996; Lawton & Prasad, 1996; Liang *et al.*, 2002). These icosahedral reconstruction programs are heavily indebted to a set of FORTRAN routines written by R. A. Crowther at the Medical Research Council Laboratory of Molecular Biology (MRC LMB) and perform the 3D reconstruction using the Fourier–Bessel transform strategy outlined above [(2.5.6.46)–(2.5.6.49)] (Crowther, Amos *et al.*, 1970; Crowther & Amos, 1972).

The general reconstruction methods (algebraic, filtered back-projection, direct Fourier inversion) easily accommodate symmetries in the data. In major single-particle reconstruction packages, reconstruction programs are implemented such that the point-group symmetry is a parameter of the program. The symmetry operation is internally taken into account during calculation of the weighting function and it is also applied to the set of Eulerian angles assigned to each projection so the multiple copies are implicitly created and processed. This approach results in an extended time of calculations, but it is entirely general. In direct Fourier inversion algorithms the numerical inaccuracy of the symmetrization performed in Fourier space will result in nonsymmetric artifacts in real space. Thus, in *SPIDER* (Frank *et al.*, 1996) an additional real-space symmetrization is performed after the reconstruction is completed. Finally, it has to be noted that although it might be tempting to calculate a 3D reconstruction without enforcement of symmetry and to symmetrize the resulting structure, this approach is incorrect. This fact can be seen from the way weighting functions are constructed: weighting functions calculated with symmetries taken into account are not equal to the weighting function calculated for a unique set of projections and subsequently symmetrized.

2.5.7. Single-particle reconstruction

By P. A. Penczek

2.5.7.1. Formation of projection images in single-particle reconstruction

Cryo-electron microscopy (cryo-EM) in combination with the single-particle approach is a new method of structure determination for large macromolecular assemblies. Currently, resolution in the range 10 to 30 Å can be reached routinely, although in a number of pilot studies it has been possible to obtain structures at 4 to 8 Å. Theoretically, electron microscopy can yield data

exceeding atomic resolution, but the difficulties in overcoming the very low signal-to-noise ratio (SNR) and low contrast in the data, combined with the adverse effects of the contrast transfer function (CTF) of the microscope, hamper progress in fulfilling the potential of the technique. However, in recent years, cryo-EM has proven its power in the structure determination of large macromolecular assemblies and machines which are too large and complex for the more traditional techniques of structural biology, *i.e.*, X-ray crystallography and NMR spectroscopy.

Single-particle reconstruction is based on the assumption that a protein exists in solution in multiple copies of the same basic structure. Unlike in crystallography, no ordering of the structure within a crystal grid is required; the enhancement of the SNR is achieved by bringing projection images of different (but structurally identical) proteins into register and averaging them. This is why the technique is sometimes called 'crystallography without crystals'.

Within the linear weak-phase-object approximation of the image formation process in the microscope [see equation (2.5.2.43) in Section 2.5.2], 2D projections represent line integrals of the Coulomb potential of the particle under examination convoluted with the point-spread function of the microscope, s, as introduced in Section 2.5.1. In addition, we have to consider the translation $\mathbf{t}$ of the projection in the plane of micrograph, suppression of high-frequency information by the envelope function E of the microscope, and two additive noises m^{B} and m^{S}. The first one is a coloured background noise, while the second is attributed to the residual scattering by the solvent or the supporting thin layer of carbon, if used, assumed to be white and affected by the transfer function of the microscope in the same way as the imaged protein. In order to have the image formation model correspond more closely to the physical reality of data collection, we write equation (2.5.6.4) from Section 2.5.6 such that the projection operation is always realized in the z direction of the coordinate system (corresponding to the direction of propagation of the electron beam), while the molecule is rotated arbitrarily by three Eulerian angles:

$$d_n(\mathbf{x}) = s_n(\mathbf{x}) * e_n(\mathbf{x}) * \left[\int f(\mathbf{T}_n\mathbf{r})\,\mathrm{d}z + m_n^{\mathrm{S}}(\mathbf{x})\right] + m_n^{\mathrm{B}}(\mathbf{x}), \quad n = 1, \ldots, N. \tag{2.5.7.1}$$

Here $f \in R^{n^3}$ represents the three-dimensional (3D) electron density of the imaged macromolecule and $d \in R^{n^2}$ is the nth observed two-dimensional (2D) projection image. The total number of projection images N depends on the structure determination project, and can vary from a few hundred to hundreds of thousands. Further, e is the inverse Fourier transform of the envelope function, $\mathbf{x} = [x \quad y]^T$ is a vector of coordinates in the plane of projections, $\mathbf{r} = [r_x \quad r_y \quad r_z \quad 1]^T$ is a vector of coordinates associated with nth macromolecule, $\mathbf{T}$ is the 4×4 transformation matrix given by

$$\mathbf{T}(\mathbf{R}, \mathbf{t}) = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{x} \\ z \\ 1 \end{bmatrix} = \mathbf{T}\mathbf{r}, \tag{2.5.7.2}$$

with $\mathbf{t} = [t_x \quad t_y \quad 1]^T$ being the shift vector of translation of the object (and its projection) in the xy plane (translation in z is irrelevant due to the projection operation) and $\mathbf{R}(\psi, \theta, \varphi)$ is the 3×3 rotation matrix specified by three Eulerian angles. As in Section 2.5.6, two of the angles define the direction of projection $\boldsymbol{\tau}(\theta, \varphi)$, while the third angle ψ results in rotation of the projection image in the plane of the formed image xy; changing this angle does not provide any additional information about the structure f. Both types of noise are assumed to be mutually uncorrelated and independent between projection images (*i.e.*, $\langle m_i^k m_j^l \rangle_{i \neq j} = 0$; k, l = S, B) and also uncorrelated with the signal ($\langle d_i m_i^k \rangle = 0$; k = S, B). Model (2.5.7.1) is semi-empirical in that, unlike in the standard model, we have two contributions to the noise. Although in principle amorphous ice should not be affected by the CTF, so the term m^{S} should be absorbed into m^{B}, in practice the buffer in which the protein is purified is not pure water and it is possible to observe CTF effects by imaging frozen buffer alone. Moreover, if a thin support carbon is used, it will be a source of very strong CTF-affected noise also included in m^{B}.

In Fourier space, (2.5.7.1) is written by taking advantage of the central section theorem [equation (2.5.6.8) of Section 2.5.6]: the Fourier transform of a projection is extracted as a Fourier plane uv of a rotated Fourier transform of a 3D object:

$$D_n(\mathbf{u}) = \mathrm{CTF}(\mathbf{u};\, \Delta f_n, q) E_n(\mathbf{u}) \Big\{ [F(\mathbf{T}\mathbf{v})]\big|_{u_z=0} + M_n^{\mathrm{S}}(\mathbf{u}) \Big\} + M_n^{\mathrm{B}}(\mathbf{u}). \tag{2.5.7.3}$$

The capital letters denote Fourier transforms of objects appearing in (2.5.7.1) while CTF (a Fourier transform of s) depends, among other parameters that are set very accurately (such as the accelerating voltage of the microscope), on the defocus setting Δf_n and the amplitude contrast ratio $0 \leq q < 1$ that reflects the presence of the amplitude contrast that is due to the removal of widely scattered electrons [the real term in (2.5.5.14)]. For the range of frequency considered, q is assumed to be constant and the CTF is written in terms of the phase perturbation function χ [given by equation (2.5.2.33)] as

$$\begin{aligned} \mathrm{CTF}(\mathbf{u};\, \Delta f) &= [1 + 2q(q-1)]^{-1/2} \big\{ (1-q) \sin[\chi(|\mathbf{u}|;\, \Delta f)] \\ &\quad - q \cos[\chi(|\mathbf{u}|;\, \Delta f)] \big\} \\ &= \sin\big\{ \chi(|\mathbf{u}|;\, \Delta f) - \arctan[q/(1-q)] \big\}, \end{aligned} \tag{2.5.7.4}$$

where for simplicity we assumed no astigmatism. Finally, the rotationally averaged power spectrum of the observed image, calculated as the expectation value of its squared Fourier intensities (2.5.7.3), is given by

$$P_d(u) = \mathrm{CTF}^2(u)\, E^2(u) \big[P_f(u) + P_{\mathrm{S}}(u) \big] + P_{\mathrm{B}}(u), \tag{2.5.7.5}$$

where $u = |\mathbf{u}|$ is the modulus of spatial frequency.

2.5.7.2. Structure determination in single-particle reconstruction

The goal of single-particle reconstruction is to determine the 3D electron-density map f of a biological macromolecule such that its projections agree in a least-squares sense with a large number of collected 2D electron-microscopy projection images, $d_n \in R^{n^2} (n = 1, 2, \ldots, N)$, of isolated (single) particles with random and unknown orientations. Thus, we seek a least-squares solution to the problem stated by (2.5.7.1) [or, equivalently, in Fourier space, to (2.5.7.3)]. This is formally written as a nonlinear optimization problem (Yang *et al.*, 2005),

$$\begin{aligned} &\min_{\psi_n, \theta_n, \varphi_n, t_{x_n}, t_{y_n}, f, \Delta f_n, q, \ldots} L(\psi_n, \theta_n, \varphi_n, t_{x_n}, t_{y_n}, f, \Delta f_n, q, \ldots) \\ &\quad \equiv \tfrac{1}{2} \sum_{n=1}^{N} \left\| s_n(\mathbf{x}) * e_n(\mathbf{x}) * \int f(\mathbf{T}_n\mathbf{r})\,\mathrm{d}z - d_n(\mathbf{x}) \right\|^2. \end{aligned} \tag{2.5.7.6}$$

The factor of $\frac{1}{2}$ is included merely for convenience. The objective function in (2.5.7.6) is clearly nonlinear due to the coupling

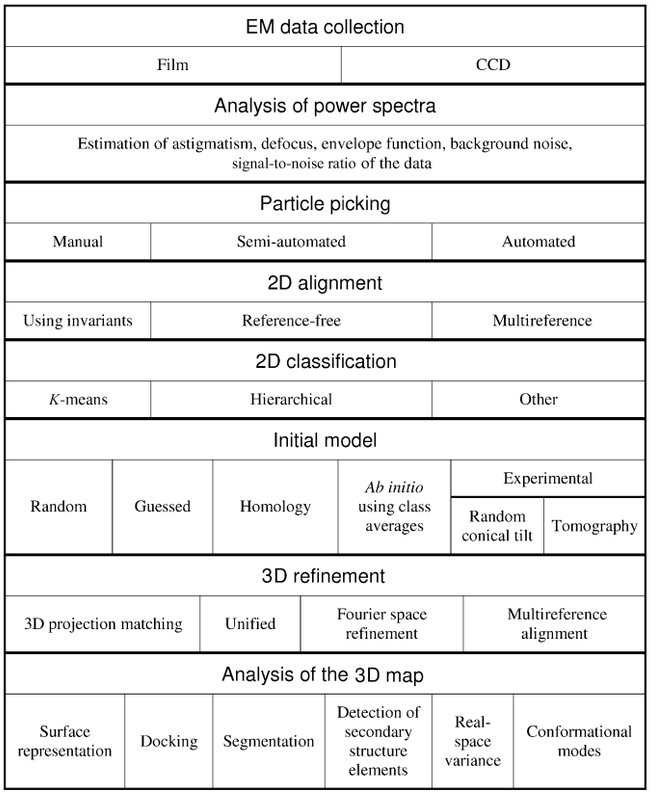


Fig. 2.5.7.1. Typical steps performed in a single-particle cryo-EM structure determination project.

between the orientation parameters $\psi_n, \theta_n, \varphi_n, t_{x_n}, t_{y_n}$ ($n = 1, 2, \ldots, N$) and the 3D density f.

The parameters in (2.5.7.6) to be determined can be separated into two groups. (1) The orientation parameters $\psi_n, \theta_n, \varphi_n, t_{x_n}, t_{y_n}$ that have to be determined entirely by solving (2.5.7.6) and for which there are no initial guesses, and the structure f itself, for which we may or may not have an initial guess. The number of parameters in this group is very large: $n^3 + 5m$. Note that in single-particle reconstruction, the number of projection data m is far greater than the linear size of the data in pixels, *i.e.*, $m \gg n$. (2) Various parameters which we will broadly call the parameters of the image formation model (2.5.7.1)–(2.5.7.4): the defocus settings of the microscope Δf_n, the amplitude contrast ratio q and, if analytical forms of the envelope function E, the power spectrum of the background noise M, or the structure F are adopted, the parameters of these equations. Some of the parameters in the second group are usually known very accurately or can be estimated from micrograph data before one attempts to solve (2.5.7.6) (see Section 2.5.7.4), but they can also be refined during the structure determination process [for the method for correcting the defocus settings, see Mouche *et al.* (2001)].

Owing to the very large number of parameters in (2.5.7.6) and the nonlinearities present, one almost never attempts to solve the problem directly. Instead, structure determination using the single-particle technique involves several steps. (i) The macromolecular complex is prepared with a purity of at least 90%. (ii) The sample is flash-frozen in liquid ethane. Alternatively, cryo-negative stain techniques or traditional negative stain methods can be used. (iii) Pictures of the macromolecular complexes are taken. (iv) Exhaustive analysis of 2D particle images aimed at increasing the SNR of the data and evaluation of the homogeneity of the sample is performed. (v) An initial low-resolution model of the structure is established using either experimental techniques or computational methods. (vi) The initial structure is refined in order to increase the resolution using an enlarged data set. Only in this step does one attempt to minimize (2.5.7.6) more or less directly. (vii) Visualization and interpretation of the resulting 3D electron-density map is the last step; it often involves docking of X-ray structures of molecules into EM density maps in order to reveal the arrangement of known molecules within the EM envelope (Fig. 2.5.7.1). As within the weak-phase-object approximation of the image formation in EM the relation between densities in collected images and the 3D electron density of the imaged macromolecule is linear [(2.5.7.1)], all data-processing methods employed in the structure determination project should be linear, so the densities in the cryo-EM 3D model can be interpreted in terms of the electron density of the protein.

In the actual single-particle project not all the steps have to be executed in the order outlined above. The technique has proved to be particularly useful in studies of functional complexes of proteins whose base state is known to a certain resolution or even of functional complexes whose atomic (X-ray crystallographic) structure is known. In these cases, steps (iv) and (v) can be omitted and the structure of the functional complex (for examples with ligands bound to it) can be relatively easily determined using the native structure as a starting point for step (vi).

In addition to difficulties with obtaining good cryo-EM data, the technique is computationally intensive. The reason is that in order to obtain a sufficient SNR in the 3D structure, processing of hundreds of thousands of EM projection images of the molecule might be necessary. For each, five orientation parameters have to be determined, and this is in addition to determination of the image-formation parameters required for the optimization of correlation searches. In effect, it is not unusual for single-particle projects to consume weeks of the computer time of multiprocessing clusters. This also explains why the knowledge of the base structure simplifies the work to a large degree: when it is known, initial values of the orientation parameters can be easily established, reducing not only the computational time, but also possibilities of errors in the structure-determination process.

2.5.7.3. Electron microscopy and data digitization

The electron microscope is a phase imaging system; *i.e.*, in order to create contrast in images, they have to be underfocused. Owing to the particular form of the CTF of the microscope [(2.5.7.4)], not only the amplitudes of the image in Fourier space are modified, but information in some ranges of spatial frequencies is set to zero and some phases have reversed sign. Therefore, in order to obtain possibly uniform coverage of Fourier space, the standard practice is to take pictures using different defocus settings and merge them computationally in order to fill gaps in Fourier space. The problem is compounded by the relation between underfocus and the envelope function of the microscope. Far-from-focus images have high contrast, but the envelope function has a relatively steep fall-off limiting the range of useful spatial frequencies. Conversely, close-to-focus images have little contrast, but the envelope function is decreasing, slowly extending useful information to high spatial frequencies. In effect, it is easier to process computationally far-from-focus data and to obtain accurate alignment of particles, but the results have severely limited resolution. Processing of close-to-focus data is challenging and results tend to be less accurate, but there is the potential to obtain high-resolution information.

The experimental techniques of initial structure determination (random conical tilt, tomography) require collection of tilt data. This is facilitated by dedicated microscope stages that can be rotated inside the microscope column yielding additional views of the same field. However, collection of high-quality tilt images is difficult. The quality of tilted images tends to be adversely affected by charging and drift effects. Moreover, as the stage is tilted the effective ice thickness increases (inversely proportionally to the cosine of the tilt angle, so at 60° the factor is two) and the contrast of the images decreases correspondingly. Finally,

the defocus in tilted micrographs varies depending on the position in the field, often forcing users to restrict the particle selection only to regions in the vicinity of the tilt axis. However, tilting establishes geometrical relations between different projections of the same particle, unambiguously allowing for robust determination of an initial 3D model and the handedness of the quaternary structure of the complex.

Electron microscope images can be either recorded on the film and subsequently converted to digital format, or they can be recorded using a charge-coupled device (CCD) camera in a digital format directly on a microscope. In either case, it is necessary to select the magnification of the microscope and the eventual pixel size of the digitized data before the data-collection session. High magnification can potentially yield high-resolution data, but at the same time it decreases the yield of particles. Lower magnification values can be used when images are recorded on film, which does not attenuate high spatial frequencies to the same extent as CCD cameras tend to do.

The pixel size has to be adjusted according to the expected resolution of the final structure. Although it is tempting to adopt a small pixel size (in the hope of achieving high resolution of the results), in most cases this is counterproductive, as it results in very large computer files that are difficult to handle and in excessively long data-processing times. Theoretically, the optimum pixel size is tied to the maximum frequency present in the data by Shannon's sampling theorem, which states that no information is lost if the signal is sampled at twice the maximum frequency present in the signal, and no additional information is gained by sampling using higher frequency. Thus, if the expected resolution is 12 Å, it should be sufficient to use a pixel size (on the specimen scale) of 6 Å. In practice, various image-processing operations performed during alignment of the data and 3D reconstruction of the complex significantly lower the range of useful frequencies. This is because in currently available single-particle reconstruction software packages rather unsophisticated interpolation schemes are employed, which were selected mainly for the speed of calculations. Therefore, it is advisable to oversample the data by a factor of 1.5 or even 3.0. For an expected resolution of 12 Å this corresponds to pixel sizes of 4 and 2 Å, respectively.

The windowed particles have to be normalized to adjust the image densities to a common framework of reference. The reason for this step is that microscopy conditions are never exactly the same and also within the same micrograph field the background densities can vary by a significant margin due to uneven ice thickness and other factors. A sensible approach to normalization is to assume that the statistical distribution of noise in areas surrounding particles should be the same (Boisset *et al.*, 1993). Hence a large portion of one of the micrographs from the processed set is selected and a reference histogram of its pixel values is generated. Next, assuming a linear transformation of pixel values, the two parameters of this transformation are found in such a way that the histogram of the transformed pixel values surrounding the particle optimally matches the reference histogram using χ^2 statistics as a discrepancy measure.

2.5.7.4. Assessment of the data quality and estimation of the image formation parameters

The initial assessment of the quality of the micrographs is usually performed during the data collection and in most cases before the micrographs are digitized. The micrographs are examined visually and those that have noticeable drift, astigmatisms, noticeable contamination or simply too low a number of particles to justify further analysis are simply discarded. After digitization of the accepted micrographs, the first step is estimation of the power spectrum, which will be examined for the presence of Thon rings (thus confirming that the micrograph is indeed usable) and astigmatism.

The method of averaged overlapping periodograms (Welch, 1967) is commonly used in EM to calculate the power spectrum. It is designed to improve the statistical properties of the estimate by taking advantage of the fact that when K identically distributed independent measurements are averaged, the variance of the average is decreased with respect to the individual variance by the ratio $1/K$. Thus, instead of calculating a periodogram (squared moduli of the discrete Fourier transform) of the entire micrograph field, one subdivides it into much smaller windows, calculates their periodograms and averages them. Typically, one would chose a window size of 512×512 pixels and an overlap of 50%, which will result in the reduction of the variance of the estimate to few percent with respect to the variance of the periodogram of the entire field (Fernandez *et al.*, 1997; Zhu *et al.*, 1997). Further reduction of the variance is achieved by rotational averaging of the 2D power-spectrum estimate. The resulting one-dimensional (1D) profile is finally used in the third step of our procedure.

For a set of micrographs the power spectra can be evaluated either visually or computationally in an automated fashion. Of main concern are the presence of Thon rings, the astigmatism and the extent to which Thon rings can be detected. Although in principle astigmatic data could be used in subsequent analysis (in fact, astigmatism could be considered advantageous, as particles from the same micrograph would contain complementary information in Fourier space), in practice they are discarded as currently there is no software that can process astigmatic data efficiently. The extent of Thon rings indicates the 'resolution' of the data, *i.e.*, the maximum frequency to which information in the data can be present.

A number of well established programs can assist the user in the calculation of power spectra and automated estimation of defocus and astigmatism (Huang *et al.*, 2003; Mindell & Grigorieff, 2003; Sander *et al.*, 2003; Mallick *et al.*, 2005). Given the analytical form of the CTF [(2.5.7.4)], the problem is solved by a robust fitting of the CTF parameters such that the analytical form of the CTF matches the power spectrum of the micrograph. Usually, the steps employed are: (1) robust estimation of the power spectrum; (2) calculation of the rotational average of the power spectrum; (3) subtraction from this rotational average of the slowly decreasing background [roughly corresponding to $P_{\rm B}$ in (2.5.7.5)]; (4) fitting of the defocus value Δf_n using known settings of the microscope (voltage, spherical aberration constant, ...) and usually assuming a constant and known value of the amplitude contrast ratio q (for cryo-EM data, q should be in the range 0.02–0.10); and (5) using the established defocus value Δf_n, analysis of the 2D power spectrum and fitting of the astigmatism amplitude and angle while refining the defocus. As long as the defocus value is not too small and there are at least two detectable zeros of the CTF, all available programs give very good and comparable results.

In some single-particle packages, the automated calculation of defocus is integrated with the estimation of additional characteristics of the image-formation parameters that are required for advanced application of a Wiener filter [(2.5.7.18)] (Saad *et al.*, 2001; Huang *et al.*, 2003), *i.e.*, the power spectra of two noise distributions $P_{\rm S}$ and $P_{\rm B}$ and the envelope function of the microscope E_n for each micrograph. A possible approach is to select slowly varying functions and fit their parameters to match the estimates of $P_{\rm S}$, $P_{\rm B}$ and P_d obtained from the data. Finally, it is necessary to have a description of the 1D rotationally averaged power spectrum of the complex P_f. One possibility is to carry out X-ray solution scattering experiments (Gabashvili *et al.*, 2000; Saad *et al.*, 2001) that yield a 1D power spectrum of the complex in solution. However, these experiments require large amounts of purified sample and the accuracy of the results in terms of the overall fall-off of the power spectrum can be disputed. For the purpose of cryo-EM, a simple approximation of the protein power spectrum by analytical functions is satisfactory.

2.5.7.5. 2D data analysis – particle picking

Depending on the properties of the imaged complex and the magnification used, a single micrograph can yield from a few to thousands of individual particle projections. The first step of the data processing is identification of particle projections in micrographs and their selection. The particles have to be windowed (boxed) using a window size exceeding the particle size by a 30–50% margin. Thus, for example, in order to determine the structure of a 550 kDa complex that has a diameter of ~120 Å to 12 Å resolution, it is appropriate to choose a pixel size of 3 Å and a window size of 60 pixels.

The selection of particles is a labour-intensive process; however, the quality of selected particle projections is a major factor in the subsequent steps of analysis and the inclusion of too many imperfect images may preclude successful determination of the 3D structure. There are three possible approaches: (1) manual selection; (2) semi-automated selection; (3) fully automated selection. In the early stages of analysis, particularly when little is known about the shape of the protein and the distribution of projection views, the manual approach is preferable. The researcher displays the micrograph on a computer screen (usually preprocessed by Fourier filtration and contrast-adjusted for better visibility of the protein) and interactively identifies locations of particle views. A trained and careful operator can yield much better results than automated approaches. The main risk is in inherent bias of a human operator – there is a tendency to focus on more familiar and more easily visible particle projections, omitting less frequently appearing orientations and in effect jeopardizing successful structure determination. In semi-automated approaches, an initial step in which putative particle projections in a micrograph are chosen is performed by a computer, all candidates are windowed and the user screens a gallery of possible particles instead of the full micrograph. Algorithms that perform the initial identification of particle views range from very simple (for example a band-pass filtration of a micrograph with subsequent selection of peaks that are no closer to each other than half of the expected particle size) to sophisticated nonlinear noise-suppression methods [for details on various algorithms see the Special Issue of the *Journal of Structural Biology* (Zhu *et al.*, 2004)]. Since the human operator will be responsible for the ultimate decision, preference is given to the faster method. In most cases, semi-automated methods are implemented within a framework of a user-friendly graphical user interface that can greatly facilitate the work. Fully automated methods are currently actively under development but, curiously, even for proteins whose high-resolution structure is known, the success rate cannot match that of a human operator (Zhu *et al.*, 2004).

The automated procedures can be divided into three groups: (1) those that rely on *ad hoc* steps of denoising and contrast enhancement followed by the search for regions of known size that emerge above the background level (Adiga *et al.*, 2004); (2) those that extract orientation-independent statistical features from regions of the micrograph that may contain particles and proceed with classification (Lata *et al.*, 1995; Hall & Patwardhan, 2004); and (3) those that employ templates, *i.e.*, either class averages of particles selected from micrographs or projections of a known 3D structure of the complex (Huang & Penczek, 2004; Sigworth, 2004).

The advantage of the first two approaches is that they do not require template images, *i.e.*, since they are based on a very broadly defined description of particles (general size, shape or abstract features derived from examples of typical particles), they are applicable in cases when no 3D structure of the complex is available. Methods from the second category usually require a training session for the algorithm to construct a set of weights for the predefined features. The methods that take advantage of the availability of templates vary greatly in complexity from straightforward cross-correlation with a generic shape (a Gaussian function, a low-passed circle) (Frank & Wagenknecht, 1984) to matched filters with large number of templates and parameters derived from the image-formation model of the micrograph (CTF and envelope functions) (Huang & Penczek, 2004; Sigworth, 2004). The motivation is clear: given an ideal object and image-formation parameters, it should be only a matter of sheer computer power and user's patience to have all particles matching the template selected. In practice, the problem is much more challenging and the success rate of template-based methods does not necessarily exceed the success rate of carefully tuned *ad hoc* methods.

One of the difficulties with the application of correlation techniques to the particle-picking problem is the unevenness of micrographs, which is caused by uneven illumination by the electron beam and, to a much larger degree, by the uneven thickness of the ice layer and, when used, the supporting carbon. A possible remedy is to calculate a 'locally normalized' cross-correlation function, in which the total variance of the micrograph is replaced by the local variance of the micrograph calculated within a window of n pixels centred on the current location l. This method has a fast implementation in Fourier space (van Heel, 1982; Roseman, 2003). A faster method is to just apply a high-pass filtration of the micrograph using a high-pass Gaussian Fourier filter with a half-width $(1/np)$ Å^{-1}, where p is the pixel size. This simple step will all but eliminate the unevenness of the micrograph background.

The main difficulty with the correlation technique is the computational complexity of the problem arising from the very large number of templates that have to be considered. The particles in the micrograph are projections of a 3D object with arbitrary in-plane rotations. In effect, to perform an exhaustive search, it is necessary to sample quasi-uniformly three Eulerian angles [equation (2.5.7.17) with $\Delta\psi = \delta\theta$]. For example, a very crude angular step of $\delta\theta = 10^\circ$ results in ~13 000 2D templates! A reduction in the number of templates can be achieved either using clustering techniques (Huang & Penczek, 2004; Wong *et al.*, 2004) or by exploring the eigenstructure of the whole set of templates (Sigworth, 2004).

2.5.7.6. 2D alignment of EM images

Alignment of pairs of 2D images is a fundamental step in single-particle reconstruction. It is aimed at bringing into register various particle projections by determining three orientation parameters (rotation angles and x and y translations) and is employed in 2D alignment of large sets of 2D noisy data and in 3D structure-refinement algorithms. The computational efficiency and numerical accuracy of this step are deciding factors in achieving high-quality structural results in an acceptable time.

All 2D alignment methods considered are aimed at finding transformation parameters such that the least-squares discrepancy between two images f and g is minimized,

$$\int \left|f(\mathbf{x}) - g(\mathbf{T}\mathbf{x})\right|^2 \mathrm{d}\mathbf{x} \rightarrow \min, \tag{2.5.7.7}$$

where $\mathbf{x} = [x \quad y \quad 1]^T$ is a vector containing the coordinates. $\mathbf{T}$ is the transformation matrix given by

$$\mathbf{T}(\alpha, x, y) = \begin{bmatrix} \cos\alpha & -\sin\alpha & t_x \\ \sin\alpha & \cos\alpha & t_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.5.7.8}$$

and is dependent on three transformation parameters: rotation angle α and two translations t_x and t_y . It has to be noted that a minimum of (2.5.7.7) can be found rapidly using the fast Fourier

transform (FFT) algorithm if only the *xy* translation is sought (2D FFT), or if only the rotation angle is needed (1D FFT).

2D alignment methods can be divided into three classes: (1) those that employ exhaustive searches in order to find three orientation parameters; (2) those that perform exhaustive searches by using either simplifications (separate searches for translation and rotation parameters) (Penczek *et al.*, 1992) or by taking advantage of invariant image representations (Schatz & van Heel, 1990; Frank *et al.*, 1992 and the following discussion; Schatz & van Heel, 1992; Marabini & Carazo, 1996); or finally (3) those that are aimed at improvement of previously determined parameters and employ local searches.

In practice, as the windowed particles are approximately centred, the search for translation parameters can be restricted to relatively small values. A very efficient algorithm that takes advantage of the geometry is based on resampling to polar coordinates of the area of the image that roughly corresponds to the particle size. The resampling is done around centres placed on pixels located within a distance from the image centre that corresponds to a preset maximum translation (Joyeux & Penczek, 2002) (Fig. 2.5.7.2). For each translation, a 1D rotational cross-correlation function in polar coordinates is calculated. Overall, the alignment method based on resampling to polar coordinates comprises the following steps: (1) the image is resampled to polar coordinates; (2) 1D FFTs of various lengths are calculated, appropriately weighted and padded with zeros to equalize their lengths; (3) complex multiplications with 1D Fourier transforms of the similarly processed referenced image are calculated; (4) the inverse 1D FFT is calculated and the position of the maximum is found. The last step yields the rotation angle. Steps (1)–(4) are repeated with the image that is being aligned shifted to account for translations. In addition, the rotation angle for one of the images being mirrored is efficiently calculated in parallel with step (3) by repeating the multiplication with the 1D Fourier transforms of the reference image complex conjugated. This additional check is a necessity in the analysis of single-particle data sets, as usually one can expect on average half of the images to be mirrored versions of the other half in the data set. Overall, the method is very accurate, because only data under the circular mask enter the calculation.

For a set of N images containing the same object in various orientations and corrupted by an additive noise, the problem of alignment would be relatively simple. For proteins that have strong preferred orientation and particularly when a staining technique is used for grid preparation, this is certainly the case. In the procedure called *reference-based alignment*, one of the images that appears 'typical' is selected and used as a reference to align the remaining images. After all available images are aligned their average is calculated and used as a reference in a repeated alignment of all images. The process is iterated until the orientations of the images stabilize (Frank *et al.*, 1982).

More formally, Frank *et al.* (1988) proposed the definition of a set of N images f_k , $k = 1, \ldots, N$, aligned if a set of transformations $\mathbf{T}_k$, $k = 1, \ldots, N$, (rotation angles and translations) is found such that all pairs of images are mutually brought into register, so the expression

$$L_1(\{f\}, \{\mathbf{T}\}) = \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} \left\| f_k(\mathbf{T}_k\mathbf{x}) - f_l(\mathbf{T}_l\mathbf{x}) \right\|^2 = \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} \left(\left\| f_k(\mathbf{T}_k\mathbf{x}) \right\|^2 + \left\| f_l(\mathbf{T}_l\mathbf{x}) \right\|^2 - 2 f_k(\mathbf{T}_k\mathbf{x}) f_l(\mathbf{T}_l\mathbf{x}) \right) \quad (2.5.7.9)$$

is minimized. Although there is no simple way to minimize L_1, the interesting observation is that there is no requirement of the images to represent the same particle, not even a similar one. This leads to the conclusion that if the minimum of L_1 could be found, a set of diverse images could be aligned; moreover, upon alignment similar images would have similar orientation and subsequent classification of such an aligned data set would reveal subsets of similar images.

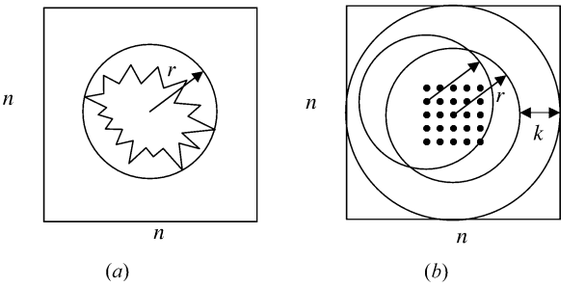


Fig. 2.5.7.2. The geometrical constraints of the 2D alignment problem. (*a*) The reference 2D particle is placed within a square image frame $n \times n$ pixels and its size is such that it can be bounded by a circle with a radius r no larger than $0.9n$. (*b*) The particle projection, the size of which is bounded by the same radius as the reference view, can be located within a circle centred on discrete locations within the image frame, such that the maximum translation is $k = (n/2) - r$. The number of possible translations is $(2k + 1)^2$. Reprinted from Joyeux & Penczek (2002) with permission from Elsevier.

A practical method of minimizing, called a *reference-free alignment*, was proposed by Penczek *et al.* (1992) by showing that minimization of L_1 is equivalent to maximization of

$$L_2(\{f\}, \{\mathbf{T}\}) = \sum_{k=1}^{N-1} \left\| f_k(\mathbf{T}_k\mathbf{x}) - \langle f \rangle_k \right\|^2, \quad (2.5.7.10)$$

where

$$\langle f \rangle_k = \frac{1}{N-1} \sum_{l=1, l\neq k}^{N} f_l(\mathbf{T}_l\mathbf{x}) \quad (2.5.7.11)$$

is the partial average of the set of images calculated with the exclusion of the kth image. The method is based on the observation that given a set of approximately aligned images, it should be possible to minimize L_2 by sequentially correcting alignments of individual images using the cross-correlation function between each image and the average of the remaining ones. On each step, depending whether the orientation of the image changes or not, (2.5.7.10) will decrease or remain constant.

The outcome of the reference-free alignment algorithm is an aligned set of N images, so all particles that have similar shapes will have similar orientations. Thus, it is natural (and because of the alignment possible) to divide the data set into classes of images that have similar shapes and orientations, *i.e.*, to cluster them. A number of well known clustering algorithms have been adopted for EM applications (Frank, 1990). The general purpose of clustering is to organize objects (in the case of EM, images) into classes whose members are similar to each other, while dissimilar to objects from other classes.

Reference-free alignment with subsequent clustering works well as long as all particles share the same overall shape (*i.e.*, the very low frequency component), as is the case for ribosomes. However, some molecules yield projections that have quite different shapes, as for example is the case for barrel-like proteins GroEL (Roseman *et al.*, 1996) with rectangular views and circular end views or flat and rectangular hemocyanin (Boisset *et al.*, 1995). In this case, the reference-free alignment tends to be unstable, as (2.5.7.10) has multiple local minima, which in practice means that the global average of the whole data set can vary significantly depending on the initiation of the procedure. In general, reference-free alignment is an 'alignment first, classifi-

cation second' approach. It is possible to reverse this order by using invariants with the supporting rationale that once approximately homogeneous classes of images were found, it should be easy to align them subsequently as within each class they will share the same motif.

A practical approach to reference-free alignment known as *alignment by classification* (Dube *et al.*, 1993) is based on the observation that for a very large data set and centred particles one can expect that although the in-plane rotation is arbitrary, there is a high chance that at least some of the similar images will be in the same rotational orientation. Therefore, in this approach the images are first (approximately) centred, then subjected to classification, and subsequently aligned.

In its simplest form, the *multireference alignment* belongs to the class of *supervised classification* methods: given a set of templates (*i.e.*, reference images; these can be selected unprocessed particle projections, or class averages that resulted from preceding analysis, or projections of a previously determined EM structure, or projections of an X-ray crystallographic structure), each of the images from the available data sets is compared (using a selected discrepancy measure) with all templates and assigned to the class represented by the most similar one. Equally often multireference alignment is understood as a form of *unsupervised classification*, more precisely *K*-means classification, even if the description is not formalized in terms of the latter. Given a number of initial 2D templates, the images are compared with all templates and assigned to the most similar one. New templates are calculated by averaging images assigned to their predecessors and the whole procedure is repeated until a stable solution is reached.

2.5.7.7. Initial determination of 3D structure using tilt experiments

The 2D analysis of projection images provides insight into the behaviour of the protein on the grid in terms of the structural consistency and the number and shape of projection images. In order to obtain 3D information, it is necessary to find geometrical relations between different observed 2D images. The most robust and historically the earliest approach is based on tilt experiments. By tilting the stage in the microscope and acquiring additional pictures of the same area of the grid it is possible to collect projection images of the same molecule with some of the required Eulerian angles determined accurately by the setting of the goniometer of the microscope.

In random conical tilt (RCT) reconstruction (Radermacher *et al.*, 1987), two micrographs of the same specimen area are collected: the first one is recorded at a tilt angle of ~50° while the second one is recorded at 0° (Fig. 2.5.7.3). If particles have preferred orientation on the support carbon film (or within the amorphous ice layer, if no carbon support is used), the projections of particles in the tilted micrographs form a conical tilt series. Since in-plane rotations of particles are random, the azimuthal angles of the projections of tilted particles are also randomly distributed; hence the name of the method. The untilted image is required for two reasons: (i) the particle projections from the untilted image are classified, thus a subset corresponding to possibly identical images can be selected ensuring that the projections originated from similar and similarly oriented structures; and (ii) the in-plane rotation angle found during alignment corresponds to the azimuthal angles in three dimensions (one of the three Eulerian angles needed). The second Eulerian angle, the tilt, is either taken from the microscope setting of the goniometer or calculated based on geometrical relations between tilted and untilted micrographs. The third Eulerian angle corresponds to the angle of the tilt axis of the microscope stage and is also calculated using the geometrical relations between two micrographs. In addition, it is necessary to centre the particle projections selected from tilted micrographs; although various correlation-based schemes have been proposed,

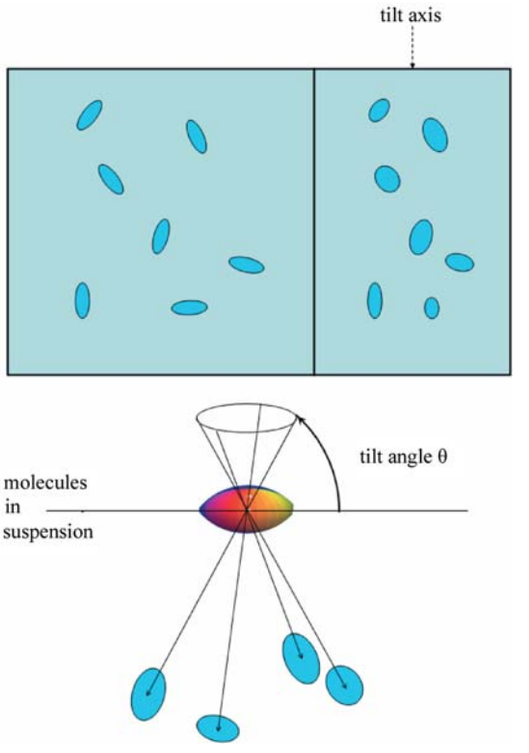


Fig. 2.5.7.3. Principle of random conical tilt reconstruction. A tilt pair of images of the same grid area is collected. By aligning the particle images in the untilted micrograph (left), the Eulerian angles of their counterparts in the tilted micrograph (right) are established. The particle images from the tilted micrograph are used for 3D reconstruction of the molecule (bottom). The set of projections form a cone in Fourier space; information within the cone remains undetermined.

the problem is difficult as the tilt data tend to be very noisy and have very low contrast.

Given three Eulerian angles and centred tilted projections, a 3D reconstruction is calculated. There are numerous advantages of the RCT method. (i) Assuming the sign of the tilt angle is read correctly (it can be confirmed by analysing the defocus gradient in the tilted micrographs), the method yields a correct hand of the structure. (ii) With the exception of the in-plane rotation of untilted projections, which can be found relatively easily using alignment procedures, the remaining parameters are determined by the experimental settings. Even if they are not extremely accurate, the possibility of a gross error is eliminated, which positively distinguishes the method from the *ab initio* computational approaches that use only untilted data. (iii) The computational analysis is entirely done using the untilted data, which have high contrast. (iv) The RCT method is often the only method of obtaining 3D information if the molecule has strongly preferential orientation and only one view is observed in untilted micrographs. The main disadvantage is that the conical projection series leaves a significant portion of the Fourier space undetermined. This follows from the central section theorem [equation (2.5.6.8) of Section 2.5.6]: as the tilt angle is less than 90°, the undetermined region can be thought to form a cone in three dimensions and is referred to as the missing cone. The problem can be overcome if the molecule has more than one preferred orientation. Subsets of particles that have similar untilted appearance (as determined by clustering) are processed independently and for each a separate 3D structure is calculated. If the preferred orientations are sufficiently different, *i.e.*, the orientations of the original particles in three dimensions are sufficiently different in terms of their angles with respect to the z axis, the 3D structures can be aligned and merged, all but eliminating the problem of the missing cone and yielding a robust, if resolution-limited, initial model of the molecule (Penczek *et al.*,

1994). It should be noted that RCT by itself almost never results in a high-resolution 3D model of the molecule. This is due to a variety of reasons, the main ones being the already mentioned poor quality of high-tilt data and difficulties with the collection of large numbers of high-quality tilted micrographs (they are often marred by drift).

In cases when the molecule does not have well defined preferred orientations, it is possible to use electron tomography to obtain the initial model. In this method, a single-axis tilt series of projection images of the same specimen area is collected using an angular step of ~2° and a maximum tilt angle not exceeding 60° (Crowther, DeRosier & Klug, 1970). The single-axis tilt data-collection geometry yields worse coverage of the Fourier space than the RCT method, leaving missing wedges uncovered (Penczek & Frank, 2006). This results in severe artifacts in real space, which make smaller objects virtually unrecognizable. The situation can be largely rectified using so-called double-axis tomography, in which a second single-axis tilt series of data are collected after rotating the specimen grid in-plane by 90° (Penczek *et al.*, 1995). This reduces the undetermined region to a missing pyramid and makes the resolution almost isotropic in the *xy* plane.

The tomographic projection data have to be aligned. This is done using either correlation techniques that enforce pairwise alignment of images (Frank & Mcewen, 1992) or by taking advantage of fiducial markers and enforcing their consistency with respect to a 3D model (Lawrence, 1992; Penczek *et al.*, 1995). In the application to single-particle work, it is possible to use locations of protein in the micrographs as markers. After the 3D reconstruction is calculated, regions collecting individual molecules are windowed from the volume and all molecules are aligned in three dimensions (Walz *et al.*, 1997). While generally robust, the procedure is labour- and computer-intensive. Unlike RCT, where only two exposures of the same field are required, electron tomography may require over one hundred images, raising serious concerns about radiation damage. Moreover, most of the data have to be collected at high tilt angle, thus are of lower quality. Particularly troublesome is alignment of 3D molecules deteriorated by the missing wedge/pyramid artifacts, with the directions of artifacts different for each object. However, when successful, electron tomography yields a very good initial model of the molecule, free from missing-Fourier-space-related artifacts and with defined handedness.

2.5.7.8. Ab initio 3D structure determination using computational methods

The experiment-based methods of initial 3D structure determination (RCT and electron tomography) are quite powerful, but rather challenging to employ in practice. Particularly frustrating is the fact that a large volume of difficult-to-record tilt data have to be collected, even though they cannot be used for subsequent high-resolution work. Therefore, whenever possible, preference is given to computational methods in which 3D geometrical relations between particle projections are established using various mathematical approaches using only untilted data.

The most straightforward approach and historically the earliest is based on the central section theorem [equation (2.5.6.8) of Section 2.5.6]: because Fourier transforms of 2D projections of a 3D object are the central section of the 3D Fourier transform of this object, it is a straightforward consequence that Fourier transforms of any two projections intersect along a line, henceforth called a *common line*. (Two trivial exceptions are the case of projections in the same direction, in which their Fourier transforms coincide with possible differences in in-plane rotation, and the case of projections in opposite directions, in which they are mirror versions of each other.) This fact was originally used by Crowther, DeRosier & Klug (1970) to solve the structure of viruses with icosahedral (60-fold) symmetry. In this case, the Fourier transform of each projection intersects itself (or rather the symmetry-related copies of itself) 37 times with the exception of degenerate cases of projections in directions of one of three symmetry axes, in which cases the number of common lines is less. Thus, it is possible to find the orientation of a single projection with respect to the chosen system of symmetry axes.

For asymmetric objects a set of three projections that do not intersect along the same line (which would correspond to the single-axis tilt geometry) uniquely determine their respective orientations (with the exception of the overall rotation, which remains arbitrary, and the handedness of the solution, which remains undetermined). Indeed, three projections span three common lines, and each common line yields two angles: for each of the intersecting sections it is the angle between the *x* axis in the system of coordinates of this section and the common line in the plane of this section. Thus, we have a total of six angles. At the same time, by arbitrarily setting the orientation of the first projection in 3D space to three Eulerian angles equal to zero (or the corresponding rotation matrix $\mathbf{R}_1 = \mathbf{I}$), we need to determine two rotation matrices $\mathbf{R}_2$ and $\mathbf{R}_3$ (or two sets of three Eulerian angles) for the remaining two projections, respectively. So, given six in-plane angles we have to find a solution for six Eulerian angles. Let the angle of the common line between the *i*th and *j*th projection in the plane of the *i*th projection be α_{ij} and the corresponding unit vector in the plane of *i*th projection be

$$\mathbf{n}_{ij} = [\cos\alpha_{ij} \quad -\sin\alpha_{ij} \quad 0]^T, \quad i, j = 1, 2, 3, \quad i \neq j, \tag{2.5.7.12}$$

where we added the third coordinate for convenience. The orientations of unit vectors $\mathbf{n}_{ij}$ in 3D space have to be related by rotation matrices; for example, vector $\mathbf{n}_{21}$ (the direction of the common line between the first and second projection in the plane of the second projection) should coincide with vector $\mathbf{n}_{12}$ (the direction of the same common line, but in the plane of the first projection) upon rotation by (unknown) matrix $\mathbf{R}_2$. All possible relations are

$$\begin{aligned} \mathbf{R}_2\mathbf{n}_{21} &= \mathbf{n}_{12} \\ \mathbf{R}_2\mathbf{n}_{23} &= \mathbf{R}_3\mathbf{n}_{32} \\ \mathbf{R}_3\mathbf{n}_{31} &= \mathbf{n}_{13}. \end{aligned} \tag{2.5.7.13}$$

Equations (2.5.7.13) have two solutions corresponding to two different hands of the molecule [for details see Farrow & Ottensmeyer (1992)].

The common-lines method works very well in the absence of noise. However, even a modest amount of noise can yield quite erroneous results or no results at all. The reason is that the solution of (2.5.7.13) is highly nonlinear with respect to the six given angles α_{ij} and small errors in the location of peaks in cross-correlation functions can lead to quite large discrepancies from the correct solution. The main difficulty with the application of the common-lines method is that the analytical solution in the form of (2.5.7.13) exists only for three projections (Goncharov *et al.*, 1987).

As a working approach to *ab initio* structure determination, the common-lines method has been implemented under the name of *angular reconstitution* in *IMAGIC* (van Heel, 1987*a*; van Heel *et al.*, 1996). In order to reduce sensitivity to noise, the method is applied not to individual projection images, but to class averages resulting from the multireference alignment of input data (van Heel *et al.*, 2000). In order to overcome the problem of the lack of solution for the larger than three number of projections, the user has to begin with selection of three judiciously chosen class averages, obtain the solution using (2.5.7.13) and subsequently include (angle) additional class averages using a brute-force approach, in which the Eulerian angles of the new projection are

calculated using a similarity measure based on common lines with the already-angled set serving as a reference.

Some of the disadvantages of the angular reconstitution were addressed in the common-lines-based method for determining orientations for $N > 3$ particle projections simultaneously (Penczek *et al.*, 1996). In this method, the problem is formulated in terms of minimization of the variance of the 3D structure, as expressed in terms of common-lines discrepancy between N projections. In a sense, the design of the method is the exact opposite of the 'standard' common-lines approach: instead of trying to the determine the Eulerian angles (rotation matrices $\mathbf{R}_i$) based on angles α_{ij} of common lines in the planes of the projections, one assumes that rotation matrices $\mathbf{R}_i$ are known, finds the set of angles α_{ij} of common lines and computes the overall discrepancy along these lines. For a pair of projections i and j, the in-plane angles of common lines are found by solving the system of equations

$$\mathbf{R}_i \mathbf{n}_{ij} = \mathbf{R}_j \mathbf{n}_{ji} \tag{2.5.7.14}$$

for α_{ij} and α_{ji}. The discrepancy minimized in the method is the variance of the 3D structure that, by analogy to the 2D case (2.5.7.10) and (2.5.7.11) is

$$L_{\rm cl}(\{F\}, \{\mathbf{R}\}) = \sum_{m=0}^{M} \sum_{k=1}^{N} \left\| F_k(u_m, \alpha; \mathbf{R}_k) - \langle F \rangle_k \right\|^2 u_m^2 \Delta u \Delta \Omega_{kl},$$
$$\langle F \rangle_k = \frac{1}{N-1} \sum_{l=1, l\neq k}^{N} F_l(u_m, \alpha; \mathbf{R}_l), \tag{2.5.7.15}$$

where $L_{\rm cl}$ is written in Fourier 3D polar coordinates and $F_k(u_m, \alpha; \mathbf{R}_k)$ is the Fourier transform of the kth projection in 2D polar coordinates (u_m, α) with the orientation in 3D Fourier space given by the rotation matrix $\mathbf{R}_k$. All Fourier planes F_k are considered to have zero thickness, so all discrepancies are calculated only along common lines and the 'partial average' $\langle F \rangle_k$ is in fact an arrangement of $N-1$ Fourier planes in 3D space. An approximation to $u_m^2 \Delta u \Delta \Omega_{kl}$ is calculated by equating the values of $\Delta \Omega_{kl}$ to the areas of the Voronoi diagram cells constructed on a unit sphere for points of intersection of common lines with this unit sphere (see Section 2.5.6.6). Generally the method performs very well, particularly if the projection images cover 3D angular space evenly.

Some macromolecules, particularly those that have an elongated barrel-like shape, will have a strongly preferential orientation with respect to the direction of electron beam showing only what are often called 'side views', *i.e.*, projections perpendicular to rotation along one axis corresponding to single-axis tilt data-collection geometry. These orthoaxial projections form single-axis tilt reconstruction geometry. In this case, Fourier transforms of all projections share only one common line, the line coinciding with the rotation axis, and clearly the common-lines-based method is not applicable to the *ab initio* structure determination. To cope with this situation, a method termed *Sidewinder* was developed (Pullan *et al.*, 2006). It is based on the observation that a Fourier transform of a finite object with a diameter D can be considered to have a nonzero thickness $1/D$ (Fig. 2.5.6.5). Thus, if the angle between two central sections of the 3D Fourier transform of this object, as derived from 2D Fourier transforms of its projections, is not too large, then these two sections will share information in Fourier space that is proportional to the amount of overlap of the two 'slabs' in Fourier space. Using this observation, the general idea employed in *Sidewinder* is to calculate pairwise cross-correlation coefficients (CCCs) between class averages of side views and to use this information to deduce the values of the azimuthal Eulerian angles using the Monte Carlo minimization method (Fishman, 1995).

For structures that have reasonably high symmetry and for those for which it is possible to collect high-quality EM data, it is sometimes possible to determine the initial structure using the 3D projection alignment method, which will be described in the next section. However, the approach is extremely computationally intensive and it is virtually impossible to try the method repeatedly to verify that the approach converges to more-or-less the same 3D structure, as is recommended for other *ab initio* methods described in this section. When the method is successful it is quite powerful, as an intermediate resolution structure can be obtained without going through intermediate and quite laborious steps of analysis of the data. A word of caution is warranted: with the direct method, unless there is external evidence that the obtained structure is correct, it is possible to obtain a self-consistent but entirely incorrect model of the molecule.

In the absence of reliable objective measures of the correctness of the structure, one can apply common sense in order to spot definitely improbable 3D maps. Given the mass of the complex it is possible to calculate the corresponding volume, and thus the threshold at which the map should be examined (Section 2.5.7.11). If at this threshold the mass density is discontinuous or there are pieces of mass surrounding the structures, the map is most likely to be incorrect. Similarly, strong directional artifacts appearing as streaks permeating the structure indicate that either the collected projection images are dominated by one or two views of the structure or that the angular assignment is incorrect. In addition, the 3D map should be centred in the window box; although the centring is not strictly speaking a mathematical requirement for a successful reconstruction, all single-particle structure-determination software packages take advantage of the fact that for centred objects orientation searches are easier to perform. So, if the map is not centred it is a clear indication of the failure of the procedure. Finally, for symmetric structures there should be no large pieces of mass on the symmetry axes.

2.5.7.9. Refinement of a 3D structure

Given an initial low-resolution model of the 3D structure and the data set of 2D projection images of the complex that have Fourier-space information extending beyond the resolution of the model, it is possible to refine the structure such that the full extent of the resolution information in the data will be utilized. In some cases, it is also possible to use as an initial structure in the refinement procedure a structure of a homologous protein, thus avoiding the process of *ab initio* structure determination altogether. The goal of the refinement is to find such orientation parameters for each of the particle projections for which (2.5.7.6) is minimized. There exist various implementations of the structure-refinement strategy and they can be roughly divided into those that perform exhaustive searches for all five orientation parameters (two translations and three Eulerian angles per 2D projection image) and those that perform local searches, usually by employing gradient information. Finally, the strategies may differ in how the correction for the CTF is implemented.

The original 3D projection-matching strategy (Penczek *et al.*, 1994) is based on the observation that given an ideal structure f and the necessary parameters of the CTF and image-formation model, it is straightforward to find five orientation parameters for each projection image. One begins with the determination of the sufficient angular step: assuming the structure is properly sampled at the Nyquist frequency and has a real-space radius of r voxels, the angular step is given by

$$\delta\theta \cong \arctan(1/r). \tag{2.5.7.16}$$

Next, keeping in mind that projection directions are parametrized by two Eulerian angles (φ, θ), one generates a set of projection directions quasi-uniformly distributed over half a unit

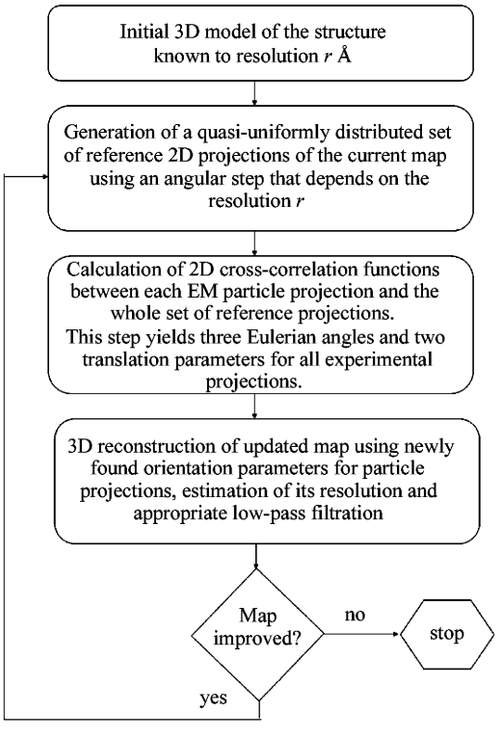

Initial 3D model of the structure known to resolution r Å
Generation of a quasi-uniformly distributed set of reference 2D projections of the current map using an angular step that depends on the resolution r
Calculation of 2D cross-correlation functions between each EM particle projection and the whole set of reference projections. This step yields three Eulerian angles and two translation parameters for all experimental projections.
3D reconstruction of updated map using newly found orientation parameters for particle projections, estimation of its resolution and appropriate low-pass filtration
Map improved?
no
stop
yes

Fig. 2.5.7.4. Schematic of the 3D projection-alignment procedure.

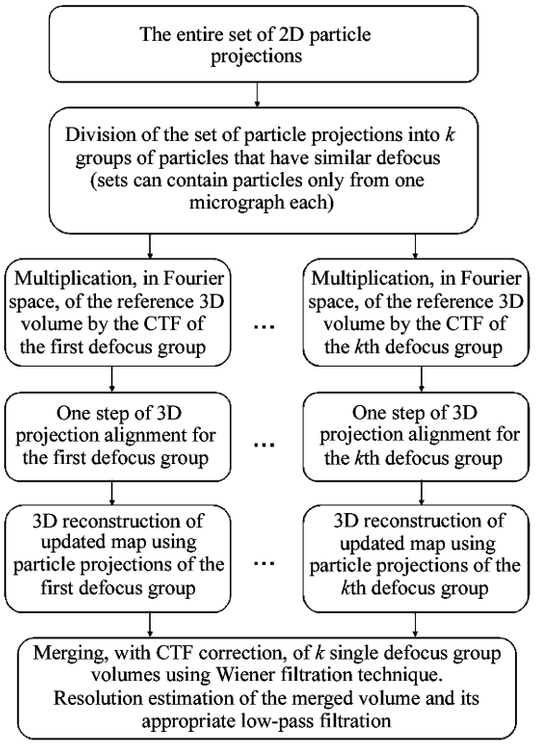

The entire set of 2D particle projections
Division of the set of particle projections into k groups of particles that have similar defocus (sets can contain particles only from one micrograph each)
Multiplication, in Fourier space, of the reference 3D volume by the CTF of the first defocus group
Multiplication, in Fourier space, of the reference 3D volume by the CTF of the kth defocus group
One step of 3D projection alignment for the first defocus group
One step of 3D projection alignment for the kth defocus group
3D reconstruction of updated map using particle projections of the first defocus group
3D reconstruction of updated map using particle projections of the kth defocus group
Merging, with CTF correction, of k single defocus group volumes using Wiener filtration technique. Resolution estimation of the merged volume and its appropriate low-pass filtration

Fig. 2.5.7.5. Schematic of 3D projection alignment with CTF correction performed on the level of 3D maps reconstructed from projection images sorted into groups that share similar defocus settings.

sphere (or, in the case of a symmetric structure, over an asymmetric subunit) by taking fixed steps along the altitude or tilt angle θ and a number of samples azimuthally in proportion to $\sin\theta$ (Penczek *et al.*, 1994). So, for a chosen constant increment $\delta\theta$ and given θ angle the increment of the φ angle varies according to

$$\Delta\varphi = \delta\theta/|\sin\theta|. \qquad (2.5.7.17)$$

If all three Eulerian angles are to be sampled, as is necessary in some applications, then ψ is sampled uniformly in steps of $\delta\theta$.

In order to find the orientation parameters of projection images, one step of projection matching is performed. The reference structure is projected in all directions given by (2.5.7.17), yielding a set of reference images. Next, for each projection image, 2D cross-correlation functions with all reference images are calculated using one of the methods described in Section 2.5.7.6 and the overall maximum yields the translation, the in-plane rotation angle, the number of the most similar reference image (thus the remaining two Eulerian angles) and information about whether the image should be mirrored. Given this, a new 3D structure can be calculated using a 3D reconstruction algorithm (see Section 2.5.6). This simple protocol constitutes the core of 3D projection alignment (Fig. 2.5.7.4).

In a simple implementation of the 3D projection-matching procedure, all projection data are assembled into *defocus groups*, *i.e.*, groups of projection images that have similar defocus settings (Frank *et al.*, 2000). During refinement, for each defocus group the reference volume is multiplied by the CTF with the appropriate defocus value, one step of projection matching is performed and a refined structure is reconstructed for this group (Fig. 2.5.7.5). In addition, the within-group resolution is estimated using the Fourier shell correlation (FSC) approach (2.5.7.19) applied to two volumes calculated from two subsets of projection images randomly split into halves. After all defocus groups have been processed, the individual refined volumes are merged in Fourier space with a CTF correction using Wiener-filter methodology (Penczek *et al.*, 1997),

$$F_{\text{merged}} = \frac{\sum_k \text{CTF}_k \text{SSNR}_k F_k}{\sum_k \text{CTF}_k^2 \text{SSNR}_k + 1}, \qquad (2.5.7.18)$$

where SSNR_k is the spectral signal-to-noise ratio estimated for each defocus group using (2.5.7.22). Subsequently, the resolution of the merged volume is estimated by merging the half-volumes into two half-merged volumes using (2.5.7.18) and comparing them using (2.5.7.19). Next, the merged volume is filtered using (2.5.7.25) and the structure is centred so that its centre of mass is placed at the centre of the volume in which it is embedded.

The 3D projection-matching approach works very well during the initial stages of the refinement as it constitutes a very efficient approach to an exhaustive search for orientation parameters of all projection data. Once the orientation parameters are known to a degree of accuracy, it is straightforward to modify the procedure such that only subsets of reference projections are generated at a time and projection images are compared only with reference projections within a specified angular distance from their angular direction established during previous iteration. This modification speeds up the procedure significantly and makes it possible to refine structures to very high resolution by using a very small angular step $\delta\theta$. Another possible modification is to introduce an additional step of 2D alignment of the projection data that share the same angular direction (Ludtke *et al.*, 1999). The advantage is that this can correct possible errors of alignment to the projection of a limited-resolution reference structure and also, to an extent, reduces the danger of bias from artifacts in the reference structure. Finally, it is also possible to incorporate into the refinement strategy a correction for the envelope function of the microscope (Ludtke *et al.*, 1999). The 3D projection-matching strategy is widely popular and most EM software packages have implementations of various versions of basic strategies, as outlined above (Frank *et al.*, 1996; Ludtke *et al.*, 1999; Hohn *et al.*, 2007).

A possible improvement over the 3D projection-matching procedure can be achieved by working in transformed spaces in which the distinction between orientation search and 3D reconstruction is removed: (1) spherical harmonics (Provencher & Vogel, 1988; Vogel & Provencher, 1988), which have found applications exclusively in the determination of icosahedral structures (Yin *et al.*, 2001, 2003); (2) Radon transform (Radermacher, 1994), with selected applications in the determination of asymmetric particles (Ruiz *et al.*, 2003); or (3) Fourier transform, implemented in the *FREALIGN* package (Grigorieff, 2007). In *FREALIGN*, the transformation between the arbitrarily oriented Fourier 2D central section and the 3D Fourier Euclidean grid is implemented using trilinear interpolation that includes *ad hoc* correction for the CTF effects. In high-resolution structure-refinement mode, the program uses a gradient-based Powell optimization algorithm (Powell, 1973), thus overcoming the main deficiency of 3D projection-matching algorithms.

A unified approach to direct minimization of (2.5.7.6) was proposed by Yang *et al.* (2005) and is implemented in the *SPARX* package as the YNP method (Hohn *et al.*, 2007). The premise of the YNP method is that the orientation parameters are approximately known (thus the initial 3D map) and both the orientation parameters and the density map are updated simultaneously in a gradient-based optimization scheme. In the YNP method, the derivatives with respect to the density distribution are calculated analytically and the derivatives with respect to orientation parameters are calculated using finite difference approximations. The YNP method is very efficient and its major advantage is that it avoids many problems associated with approximate solutions inherent in methods that work in transform spaces. The projection/backprojection operations are carried out rapidly using linear interpolation, which due to sufficient oversampling of the data does not have a significant adverse impact on the solution. Moreover, because the density map f is updated simultaneously with the orientation parameters, the computationally demanding separate step of 3D reconstruction is eliminated.

2.5.7.10. Resolution estimation and analysis of errors in single-particle reconstruction

The development of resolution measures in EM was greatly influenced by earlier work in X-ray crystallography. In EM, the problem is somewhat more difficult as, unlike in crystallography, both the amplitude and the phase information in the data are affected by alignment procedures (which we consider distant analogues of phase-extension methods in crystallography). Therefore, resolution measures in EM reflect the self-consistency of the results; however, as the data are subject to alignment, there is a significant risk of introducing artifacts resulting from the alignment of the noise component in the data. Ultimately, these artifacts will unduly 'improve' the resolution of the map.

The resolution measures used in EM fall into two categories: measures based on averaging of Fourier transforms of individual images and measures based on comparisons of averages calculated for subsets of the data. In the first group, we have the Q-factor (van Heel & Hollenberg, 1980; Kessel *et al.*, 1985) and the spectral signal-to-noise ratio (SSNR) introduced for the 2D case by Unser and co-workers (Unser *et al.*, 1987), and for the 3D case for a class of reconstruction algorithms data are based on direct Fourier inversion by Penczek (Penczek, 2002). The second group of measures includes the differential phase residual (DPR) (Frank *et al.*, 1981) and the Fourier ring correlation (FRC) (Saxton & Baumeister, 1982). A marked advantage of these measures is that they are equally well applicable to 2D or 3D data. In the latter case, the volumes resulting from 3D reconstruction algorithms take the place of the 2D averages.

The resolution measures used in single-particle reconstruction are designed to evaluate the SSNR in the reconstruction as a function of spatial frequency (Penczek, 2002). The 'resolution' of the reconstruction is reported as a spatial frequency limit beyond which the SSNR drops below a selected level, for example below one.

The FSC is evaluated by taking advantage of the large number of single-particle images: the total data set is randomly split into halves; for each subset a 3D reconstruction is calculated (in two dimensions, a simple average); and two maps f and g are compared in Fourier space,

$$\mathrm{FSC}(f, g; u) = \frac{\sum_{|\|\mathbf{u}_n\|-r|\le\varepsilon}^{n_r} F(\mathbf{u}_n)G^*(\mathbf{u}_n)}{\left\{\left[\sum_{|\|\mathbf{u}_n\|-r|\le\varepsilon}^{n_r} |F(\mathbf{u}_n)|^2\right]\left[\sum_{|\|\mathbf{u}_n\|-r|\le\varepsilon}^{n_r} |G(\mathbf{u}_n)|^2\right]\right\}^{1/2}}. \quad (2.5.7.19)$$

In (2.5.7.19), 2ε is a preselected ring/shell thickness, the $\mathbf{u}_n$ form a uniform grid in Fourier space, $u = \|\mathbf{u}_n\|$ is the magnitude of the spatial frequency and n_r is the number of Fourier voxels in the shell corresponding to frequency u. The FSC yields a 1D curve of correlation coefficients as a function of u. Note that the FSC is insensitive to linear transformations of the densities of the objects. An FSC curve everywhere close to one reflects strong similarity between f and g; an FSC curve with values close to zero indicates the lack of similarity between f and g. Particularly convenient for the interpretation of the results in terms of 'resolution' is the relation between the FSC and the SSNR, which is easily derived by taking the expectation of (2.5.7.19) under the assumption that both f and g are sums of the same signal and different realizations of the noise, which are uncorrelated with the signal and between them (Saxton, 1978):

$$E[\mathrm{FSC}] \cong \frac{\mathrm{SSNR}}{\mathrm{SSNR}+1}. \quad (2.5.7.20)$$

By solving (2.5.7.20) for SSNR we obtain

$$\mathrm{SSNR} = \frac{\mathrm{FSC}}{1-\mathrm{FSC}}, \quad (2.5.7.21)$$

which, taking into account that the FSC was calculated from the data set split into halves, has to be modified to (Unser *et al.*, 1987)

$$\mathrm{SSNR} = 2\left(\frac{\mathrm{FSC}}{1-\mathrm{FSC}}\right). \quad (2.5.7.22)$$

In order to calculate the FSC that corresponds to a given SSNR, one inverts (2.5.7.22) to

$$\mathrm{FSC} = \frac{\mathrm{SSNR}}{\mathrm{SSNR}+2}. \quad (2.5.7.23)$$

Equations (2.5.7.21) and (2.5.7.22) serve as a basis for various 'resolution criteria' used in EM. The often-used 3σ criterion (van Heel, 1987*b*) equates resolution with the point at which the FSC is larger than zero at a 3σ level, where σ is the expected standard deviation of the FSC that has an expected value of zero, in essence finding a frequency for which the SSNR is significantly larger than zero. The 3σ criterion has a distinct disadvantage of reporting the resolution at a frequency at which there is no significant signal, while tempting the user to interpret the detail in the map at this resolution. Moreover, as the FSC approaches zero, its relative error increases, so the curve oscillates widely

around the zero level increasing the chance of selecting an incorrect resolution point. In other criteria one tries to equate the resolution with the frequency at which noise begins to dominate the signal. A good choice of the cut-off level is SSNR = 1.0, a level at which the power of the signal in the reconstruction is equal to the power of the noise. According to (2.5.7.22), this corresponds to FSC = 0.333. Another often-used cut-off level is FSC = 0.5, at which the SSNR in the reconstruction is 2.0 (Böttcher *et al.*, 1997; Conway *et al.*, 1997; Penczek, 1998).

The main reason behind the determination of the resolution of the EM maps is the necessary step of low-pass filtration of the results before the interpretation of the map is attempted. In order to avoid mistakes, particularly the danger of overinterpretation, one has to remove from the map unreliable Fourier coefficients. Inclusion of Fourier coefficients with a low SNR will result in the creation of spurious details and artifacts in the map. Thus, the optimal filtration should be based on the SSNR distribution in the map and the solution is given by a Wiener filter:

$$W(u) = \frac{\mathrm{SSNR}(u)}{\mathrm{SSNR}(u)+1}. \qquad (2.5.7.24)$$

Based on the relation of FSC to SSNR (2.5.7.22), we can write (2.5.7.24) as

$$W(u) = \frac{2\mathrm{FSC}(u)}{\mathrm{FSC}(u)+1}. \qquad (2.5.7.25)$$

In practice, because of the irregular shape of typical FSC curves (particularly for small values of FSC) it is preferable to approximate the shape of the Wiener filter (2.5.7.25) by one of the standard low-pass filters, such as Butterworth (Gonzalez & Woods, 2002) or hyperbolic tangent (Basokur, 1998).

The FRC/FSC methodology can be used to compare a noise-corrupted map with a noise-free ideal version of the same object. In single-particle reconstruction this situation emerges when an X-ray crystallographic structure of either the entire EM-determined structure or of some of its domains is available (Penczek *et al.*, 1999). In this case, we assume that in (2.5.7.19) f represents a sum of the signal and additive uncorrelated noise and g represents the noise-free signal, so is straightforward to calculate the expectation of (2.5.7.19) in order to obtain the relation between the cross-resolution (CRC) and the SSNR:

$$E[\mathrm{CRC}] \cong \left(\frac{\mathrm{SSNR}}{\mathrm{SSNR}+1}\right)^{1/2}. \qquad (2.5.7.26)$$

Thus

$$\mathrm{SSNR} = \frac{\mathrm{CRC}^2}{1-\mathrm{CRC}^2}. \qquad (2.5.7.27)$$

Interestingly, for the same SSNR cut-off levels, corresponding values of CRC are higher than those for FSC. For example, for SSNR = 1, CRC = 0.71, while FSC = 0.33. For SSNR = 2, CRC = 0.82, while FSC = 0.5.

2.5.7.11. Analysis of 3D cryo-EM maps

The amount of structural information that can be derived from a structure of a macromolecular complex determined by cryo-EM depends on two factors: the resolution of the map and the availability of additional structural information about the system. Generally, we will refer to complexes at a resolution better than 7 Å as high-resolution structures, as at this resolution the elements of secondary structure become directly visible. Maps at resolution lower than that we will call intermediate resolution, as at this scale of detail one can only determine a general arrangement of subunits. However, it is good to realize that there is a huge difference between the amount and reliability of information derived from a map of the same complex determined at 10 Å as compared to a map determined at 30 Å resolution. Similarly, very large complexes determined at 50 Å resolution will yield more information than very small complexes determined at 15 Å resolution. On the other hand, even intermediate-resolution EM maps provide extremely valuable information if they can be placed in the context of other structural work. The single-particle structure can be also investigated within a context of a more complex system using other, lower-resolution techniques, for example electron tomography. In this case, by using docking approaches one can determine the distribution, orientation and general arrangement of smaller cryo-EM determined complexes within larger subcellular systems. On a different scale of resolution, it is quite common to have structures of some domains or even of the entire complexes determined to atomic resolution by X-ray crystallography. Again, by using docking techniques it is possible to determine whether the conformation of the EM structure differs from that determined by X-ray crystallography or to map subunits and domains of the larger complex by fitting available atomic resolution structures.

The basic mode of visualization of cryo-EM maps is surface representation. The first step involves the choice of an appropriate threshold level for the displayed surface, particularly when the scaling of the cryo-EM data is arbitrary. A good guide is provided by the total molecular mass of the complex: given a pixel size of p Å, an average protein density $d = 1.36 \times 10^{-24}$ g Å^{-3} and the total molecular mass of the complex M Da, the number of voxels N_v occupied by the complex is

$$N_\mathrm{v} = M/\left(p^3 d N_\mathrm{A}\right), \qquad (2.5.7.28)$$

where N_A is the Avogadro's number (6.02×10^{23} atoms mole^{-1}). Based on that, one can find the threshold that for a given structure encompasses the determined number of voxels N_v [appropriate functions are implemented in *SPIDER* (Agrawal *et al.*, 1996; Frank *et al.*, 1996) and *SPARX* (Hohn *et al.*, 2007)]. At a sufficiently high resolution, cryo-EM maps can be analysed in the same manner as X-ray crystallographic maps and using the same graphical/analytical packages ('backbone tracing') (Jones *et al.*, 1991) (Fig. 2.5.7.6).

The complexity of cryo-EM maps of large macromolecular assemblies combined with their limited resolution invites attempts to automate some of the steps of analysis in an attempt to make the results more robust and less dependent on the researcher's bias. A good example of semi-automated analysis is the nucleic acid–protein separation in a 11.5 Å cryo-EM map of the 70S *E. coli* ribosome (Spahn *et al.*, 2000). In the procedure, the (continuous-valued) densities were analysed making use of (i) the difference in scattering density between protein and nucleic acids; (ii) continuity constraints that the image of any nucleic acid molecule must obey and (iii) knowledge of the molecular volumes of all proteins. As a result, it was possible to reproduce boundary assignments between ribosomal RNA (rRNA) and proteins made from higher-resolution X-ray maps of the ribosomal subunits with a high degree of accuracy, and allowed plausible predictions to be made for the placements of proteins and RNA components as yet unassigned. One of the conclusions derived from this separation was that the 23S rRNA is solely responsible for the catalysis of peptide-bond formation; thus, the ribosome is a ribozyme. The same conclusion was reached independently in the studies of the X-ray crystallographic structure of the 70S ribosome (Nissen *et al.*, 2000). The method by Spahn *et al.* cannot be easily extended to other

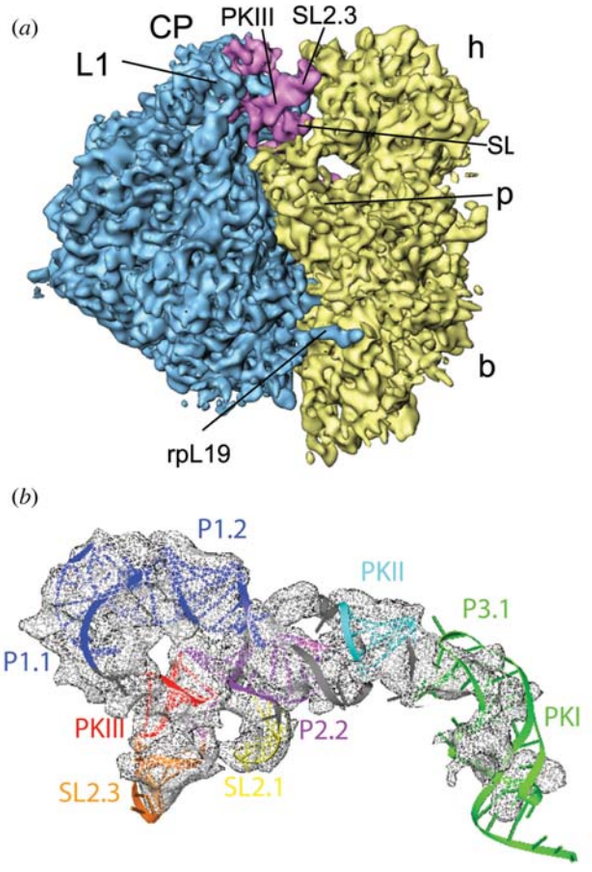


Fig. 2.5.7.6. A high-resolution cryo-EM map allows backbone tracing. (*a*) Cryo-EM map of the cricket paralysis virus (CrPV) IRES RNA in complex with the yeast 80S ribosome determined at 7.3 Å resolution. The map is shown from the L1 protuberance side with the ribosomal 40S subunit in yellow, the 60S subunit in blue and the CrPV IRES in magenta. Landmarks for the 40S subunit: b, body; h, head; p, platform. Landmarks for the 60S subunit: CP, central protuberance; L1, L1 protuberance. PKIII denotes helix PKIII of the CrPV IRES RNA and SL the two stem loops present in the secondary structure of the RNA. (*b*) Structure of the CrPV IRES RNA. Based on the cryo-EM map and additional biochemical knowledge, the complete chain of the RNA is found (189 nucleotides could be traced). The IRES molecular model is shown as a coloured ribbon docked into the cryo-EM density (grey mesh). PK and P denote the individual helical elements of the IRES and SL denotes the stem loops (Schüler *et al.*, 2006).

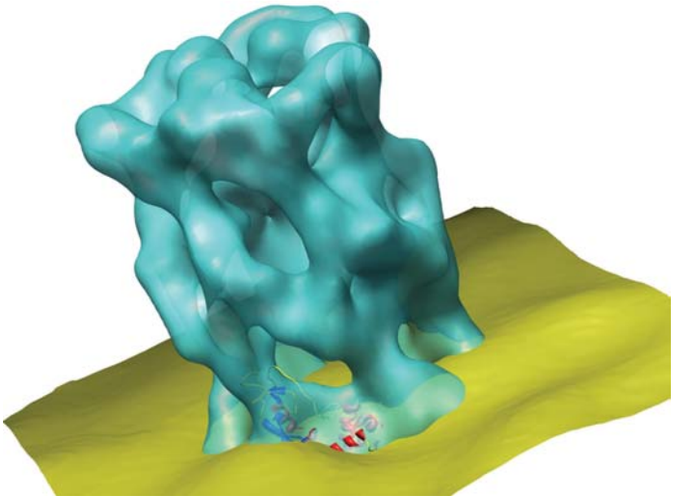


Fig. 2.5.7.7. Hrs (blue) embedded into the membrane (yellow) of an early endosome. Two functional domains have been docked into the cryo-EM density map of hexameric Hrs, determined to 16 Å resolution, and are shown as ribbons coloured by secondary structure (Pullan *et al.*, 2006). The structures of the VHS and FYVE domains have been determined crystallographically (Mao *et al.*, 2000) and are docked into the EM density map of the Hrs. The knowledge of the location of the FYVE and VHS domains, which are reported to bind to PI(3)P molecules found within the endosomal membrane (Kutateladze *et al.*, 1999), has guided the hypothetical placement of Hrs within the endosomal membrane. The immersion of the end caps of Hrs into the endosomal membrane demonstrates an 'end-on' binding model of the Hrs particles with the membrane. According to this model, either end cap can embed into the membrane, allowing the other end cap to carry out other essential protein trafficking functions, or to embed into another membrane, thus preventing fusion of membranes during early endosomal fusion.

macromolecules that comprise only protein and generally it is very difficult to delineate at intermediate resolution subunits of large macromolecular assemblies, automatically or not, in the absence of independent knowledge about their shape. The reason is that both the density and shape of the subunits are affected by the limited resolution differently depending on their spatial context. In general, subunits that are isolated and located on the surface or protruding from the structure will have relatively lower density while at the same time their overall shape will be better preserved and easier to discern. Subunits located inside the structure and surrounded by other structural elements, while having higher density, are more difficult to recognize, as they fuse with the surrounding mass densities. Therefore, it is difficult to provide a general method that could cope with the problem of automated mass-density analysis.

As most cryo-EM structures are determined at intermediate resolution, the most common mode of analysis is either to compare the map with the available X-ray crystallographic structures of its domains or to consider the result in the context of larger, subcellular structures obtained by electron tomography. In both cases correlation techniques are used extensively to obtain objective results or to validate the results obtained by manual fitting.

In docking of X-ray crystallographic structures into EM maps, the first step is the conversion of atomic coordinates from X-ray molecular models, as given in Protein Data Bank (PDB) files, into an electron-density map in a way that would mimic the physical image formation process. Although sophisticated methods of computational emulation of the image-formation process in the electron microscope are available, very simple approaches to conversion yield quite satisfactory results at the resolution of the EM results. The most common one is to assume that the Coulomb potential of an atom is proportional to its atomic number and add these atomic numbers within a Euclidean grid with a cell size equal to the EM pixel size in Å. The atomic coordinates of atoms are interpolated within the grid using trilinear interpolation. After such conversion, the X-ray map can be handled using the general image-processing tools of a single-particle software package. Initial orientation (or orientations, if the general placement is not immediately visually apparent) of the X-ray map can be easily performed manually within any number of graphical packages, for example *Chimera* (Pettersen *et al.*, 2004). The initial six orientation parameters (three translations and three Eulerian angles) are next transferred to the EM package (for details see Baldwin & Penczek, 2007) and the manual docking is refined using correlation techniques (Fig. 2.5.7.7). Similarly, the handedness of the EM map can be established or confirmed by performing fitting of the X-ray determined structure to two EM maps that differ by their hand.

Docking of EM maps into the broader cellular context of structures determined by electron tomography can provide information about the distribution of complexes and their interactions within the cell. Conceptually, the approach is very similar to that of particle picking, *i.e.*, template matching, with the main difference being that calculations are performed in three instead of two dimensions. Given a 3D structure of a single-particle EM complex, a set of 3D templates is prepared by rotating the

template around its centre of mass using the quasi-uniformly distributed three Eulerian angles [equation (2.5.7.17) with $\Delta\psi = \delta\theta$]. However, in application to tomography the angular step $\delta\theta$ can be relatively large, resulting in a much smaller number of templates than in two dimensions, the reason being the rather low resolution of typical electron tomograms (not exceeding 50 Å). Next, a brute-force 3D cross-correlation search with all templates is performed (Frangakis *et al.*, 2002). After windowing out 3D subvolumes containing putative complexes, subsequent averaging and classification can be performed.

Cryo-EM is a unique structural technique in its ability to detect conformational variability of large molecular assemblies within one sample that may contain a mixture of complexes in various conformational states. In addition to the expected conformational heterogeneity of the assemblies, due to fluctuations of the structure around the ground state one can expect to capture molecules in different functional states, especially if the binding of a ligand induces a conformational change in the macromolecular assembly. Therefore, a data set of images from an EM experiment must be interpreted as a mixture of projections from similar but not identical structures. The analysis of the extent of the resulting variability requires the calculation of the real-space distribution of 3D variance/covariance in macromolecules reconstructed from a set of their projections. The problem is difficult, as there is no clear relation between the variance in sets of projections that have the same angular direction and the variance of the 3D structure calculated from these projections. Penczek, Chao *et al.* (2006) proposed calculating the variance in the 3D mass distribution of the structure using a statistical bootstrap resampling technique, in which a new set of projections is selected with replacements from the available whole set of N projections. In the new set, some of the original projections will appear more than once, while others will be omitted. This selection process is repeated a number of times and for each new set of projections the corresponding 3D volume is calculated. Next, the voxel-by-voxel bootstrap variance $\sigma_{\rm B}^2$ of the resulting set of volumes is calculated. The target variance is obtained using a relationship between the variance of arithmetic means for sampling with replacements and the sample variance,

$$\sigma^2 = N\sigma_{\rm B}^2. \qquad (2.5.7.29)$$

The estimated structure-variance map can be used for (i) detection of different functional states (for example, those characterized by binding of a ligand) and subsequent classification of the data set into homogeneous groups (Penczek, Frank & Spahn, 2006*a*), (ii) analysis of the significance of small details in 3D reconstructions, (iii) analysis of the significance of details in difference maps, and (iv) docking of known structural domains into EM density maps.

The bootstrap technique also leads to the analysis of conformational modes of macromolecular complexes, and this is due to the fact that the covariance matrix of the structure can be directly calculated from the bootstrap volumes. The covariance matrix obtained this way would be very large. One possibility is to calculate only correlation coefficients between regions of interest that have large variance (Penczek, Chao *et al.*, 2006). Another possibility is to use the iterative Lanczos technique (Parlett, 1980) and calculate eigenvolumes directly from bootstrap volumes without forming the covariance matrix. These eigenvolumes are related to conformational modes of the molecule, as captured by the projection data of the sample (Penczek, Frank & Spahn, 2006*b*). Thus, this direct relation to the actual cryo-EM projection data positively distinguishes this approach from other techniques in which conformations are postulated based on flexible models of the EM map (Ming *et al.*, 2002; Mitra *et al.*, 2005).

2.5.8. Direct phase determination in electron crystallography

By D. L. Dorset

2.5.8.1. Problems with 'traditional' phasing techniques

The concept of using experimental electron-diffraction intensities for quantitative crystal structure analyses has already been presented in Section 2.5.4. Another aspect of quantitative structure analysis, employing high-resolution images, has been presented in Sections 2.5.5 to 2.5.7. That is to say, electron micrographs can be regarded as an independent source of crystallographic phases.

Before direct methods (Chapter 2.2) were developed as the standard technique for structure determination in small-molecule X-ray crystallography, there were two principal approaches to solving the crystallographic phase problem. First, 'trial and error' was used, finding some means to construct a reasonable model for the crystal structure *a priori*, *e.g.* by matching symmetry properties shared by the point group of the molecule or atomic cluster and the unit-cell space group. Secondly, the autocorrelation function of the crystal, known as the Patterson function (Chapter 2.3), was calculated (by the direct Fourier transform of the available intensity data) to locate salient interatomic vectors within the unit cell.

The same techniques had been used for electron-diffraction structure analysis (nowadays known as *electron crystallography*). In fact, advocacy of the first method persists. Because of the perturbations of diffracted intensities by multiple-beam dynamical scattering (Chapter 5.2), it has often been suggested that trial and error be used to construct the scattering model for the unit crystal in order to test its convergence to observed data after simulation of the scattering events through the crystal. This indirect approach assumes that no information about the crystal structure can be obtained directly from observed intensity data. Under more favourable scattering conditions nearer to the kinematical approximation, *i.e.* for experimental data from thin crystals made up of light atoms, trial-and-error modelling, simultaneously minimizing an atom–atom nonbonded potential function with the crystallographic residual, has enjoyed widespread use in electron crystallography, especially for the determination of linear polymer structures (Brisse, 1989; Pérez & Chanzy, 1989).

Interpretation of Patterson maps has also been important for structure analysis in electron crystallography. Applications have been discussed by Vainshtein (1964), Zvyagin (1967) and Dorset (1994*a*). In face of the dynamical scattering effects for electron scattering from heavy-atom crystals realized later (*e.g.* Cowley & Moodie, 1959), attempts had also been made to modify this autocorrelation function by using a power series in $|F_{\mathbf{h}}|$ to sharpen the peaks (Cowley, 1956). (Here $F_{\mathbf{h}} \equiv \Phi_{\mathbf{h}}$, replacing the notation for the kinematical electron-diffraction structure factor employed in Section 2.5.4.) More recently, Vincent and coworkers have selected first-order-Laue-zone data from inorganics to minimize the effect of dynamical scattering on the interpretability of their Patterson maps (Vincent & Exelby, 1991, 1993; Vincent & Midgley, 1994). Vainshtein & Klechkovskaya (1993) have also reported use of the Patterson function to solve the crystal structure of a lead soap from texture electron-diffraction intensity data.

It is apparent that trial-and-error techniques are most appropriate for *ab initio* structure analysis when the underlying crystal structures are reasonably easy to model. The requisite positioning of molecular (or atomic) groups within the unit cell may be facilitated by finding atoms that fit a special symmetry position [see *IT* A (2005)]. Alternatively, it is helpful to know the molecular orientation within the unit cell (*e.g.* provided by the Patterson function) to allow the model to be positioned for a conformational or translational search. [Examples would include

the polymer-structure analyses cited above, as well as the layer-packing analysis of some phospholipids (Dorset, 1987).] While attempts at *ab initio* modelling of three-dimensional crystal structures, by searching an n-dimensional parameter space and seeking a global internal energy minimum, has remained an active research area, most success so far seems to have been realized with the prediction of two-dimensional layers (Scaringe, 1992). In general, for complicated unit cells, determination of a structure by trial and error is very difficult unless adequate constraints can be placed on the search.

Although Patterson techniques have been very useful in electron crystallography, there are also inherent difficulties in their use, particularly for locating heavy atoms. As will be appreciated from comparison of scattering-factor tables for X-rays [*IT* C (2004), Chapter 6.1] with those for electrons [*IT* C (2004), Chapter 4.3], the relative values of the electron form factors are more compressed with respect to atomic number than are those for X-ray scattering. As discussed in Chapter 2.3, it is desirable that the ratio of summed scattering-factor terms, $r = \sum_{\text{heavy}} Z^2 / \sum_{\text{light}} Z^2$, where Z is the scattering-factor value at $\sin\theta/\lambda = 0$, be near unity. A practical comparison would be the value of r for copper (DL-alaninate) solved from electron-diffraction data by Vainshtein *et al.* (1971). For electron diffraction, $r = 0.47$ compared to the value 2.36 for X-ray diffraction. Orientation of salient structural features, such as chains and rings, would be equally useful for light-atom moieties in electron or X-ray crystallography with Patterson techniques. As structures become more complicated, interpretation of Patterson maps becomes more and more difficult unless an automated search can be carried out against a known structural fragment (Chapter 2.3).

2.5.8.2. Direct phase determination from electron micrographs

The 'direct method' most familiar to the electron microscopist is the high-resolution electron micrograph of a crystalline lattice. Retrieval of an average structure from such a micrograph assumes that the experimental image conforms adequately to the 'weak phase object' approximation, as discussed in Section 2.5.5. If this is so, the use of image-averaging techniques, *e.g.* Fourier filtration or correlational alignment, will allow the unit-cell contents to be visualized after the electron-microscope phase contrast transfer function is deconvoluted from the average image, also discussed in Section 2.5.5. Image analyses can also be extended to three dimensions, as discussed in Section 2.5.6, basically by employing tomographic reconstruction techniques to combine information from the several tilt projections taken from the crystalline object. The potential distribution of the unit cell to the resolution of the imaging experiment can then be used, *via* the Fourier transform, to obtain crystallographic phases for the electron-diffraction amplitudes recorded at the same resolution. This method for phase determination has been the mainstay of protein electron crystallography.

Once a set of phases is obtained from the Fourier transform of the deconvoluted image, they must, however, be referred to an allowed crystallographic origin. For many crystallographic space groups, this choice of origin may coincide with the location of a major symmetry element in the unit cell [see *IT* A (2005)]. Hence, since the Fourier transform of translation is a phase term, if an image shift $[\delta(\mathbf{r} + \mathbf{r}_0)]$ is required to translate the origin of the repeating mass unit $\varphi(\mathbf{r})$ from the arbitrary position in the image to a specific site allowed by the space group,

$$\mathrm{g}(\mathbf{r}) = \varphi(\mathbf{r}) \otimes \delta(\mathbf{r} + \mathbf{r}_0) = \varphi(\mathbf{r} + \mathbf{r}_0),$$

where the operation '$\otimes$' denotes convolution. The Fourier transform of this shifted density function will be

$$G(\mathbf{s}) = F(\mathbf{s}) \exp(2\pi i \mathbf{s} \cdot \mathbf{r}_0) = |F(\mathbf{s})| \exp[i(\phi_s + 2\pi i \mathbf{s} \cdot \mathbf{r}_0)].$$

In addition to the crystallographic phases ϕ_s, it will, therefore, be necessary to find the additional phase-shift term $2\pi i \mathbf{s} \cdot \mathbf{r}_0$ that will access an allowed unit-cell origin. Such origin searches are carried out automatically by some commercial image-averaging computer-software packages.

In addition to applications to thin protein crystals (*e.g.* Henderson *et al.*, 1990; Jap *et al.*, 1991; Kühlbrandt *et al.*, 1994), there are numerous examples of molecular crystals that have been imaged to a resolution of 3–4 Å, many of which have been discussed by Fryer (1993). For π-delocalized compounds, which are the most stable in the electron beam against radiation damage, the best results (2 Å resolution) have been obtained at 500 kV from copper perchlorophthalocyanine epitaxically crystallized onto KCl. As shown by Uyeda *et al.* (1978–1979), the averaged images clearly depict the positions of the heavy Cu and Cl atoms, while the positions of the light atoms in the organic residue are not resolved. (The utility of image-derived phases as a basis set for phase extension will be discussed below.) A number of aromatic polymer crystals have also been imaged to about 3 Å resolution, as reviewed recently (Tsuji, 1989; Dorset, 1994*b*).

Aliphatic molecular crystals are much more difficult to study because of their increased radiation sensitivity. Nevertheless, monolamellar crystals of the paraffin n-$C_{44}H_{90}$ have been imaged to 2.5 Å resolution with a liquid-helium cryomicroscope (Zemlin *et al.*, 1985). Similar images have been obtained at room temperature from polyethylene (Revol & Manley, 1986) and also a number of other aliphatic polymer crystals (Revol, 1991).

As noted by J. M. Cowley and J. C. H. Spence in Section 2.5.1, dynamical scattering can pose a significant barrier to the direct interpretation of high-resolution images from many inorganic materials. Nevertheless, with adequate control of experimental conditions (limiting crystal thickness, use of high-voltage electrons) some progress has been made. Pan & Crozier (1993) have described 2.0 Å images from zeolites in terms of the phase-grating approximation. A three-dimensional structural study has been carried out on an aluminosilicate by Wenk *et al.* (1992) with thin samples that conform to the weak-phase-object approximation at the 800 kV used for the imaging experiment. Heavy and light (*e.g.* oxygen) atoms were located in the micrographs in good agreement with an X-ray crystal structure. Heavy-atom positions from electron microscopic and X-ray structure analyses have also been favourably compared for two heavy-metal oxides (Hovmöller *et al.*, 1984; Li & Hovmöller, 1988).

2.5.8.3. Probabilistic estimate of phase invariant sums

Conventional direct phasing techniques, as commonly employed in X-ray crystallography (*e.g.* see Chapter 2.2), have also been used for *ab initio* electron-crystallographic analyses. As in X-ray crystallography, probabilistic estimates of a linear combination of phases (Hauptman & Karle, 1953; Hauptman, 1972) are made after normalized structure factors are calculated *via* electron form factors, *i.e.*

$$|E_{\mathbf{h}}^2| = I_{\text{obs}} / \varepsilon \sum_i f_i^2, \text{ where } \langle |E|^2 \rangle = 1.000.$$

(Here, an overall temperature factor can be found from a Wilson plot. Because of multiple scattering, the value of B may be found occasionally to lie close to 0.0 Å^2.) The phase invariant sums

$$\psi = \phi_{\mathbf{h}_1} + \phi_{\mathbf{h}_2} + \phi_{\mathbf{h}_3} + \ldots$$

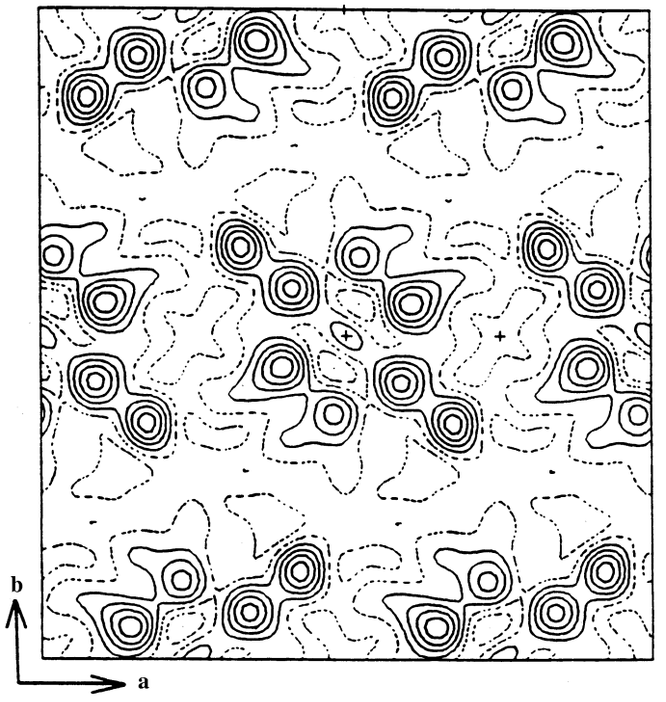


Fig. 2.5.8.1. Potential map for diketopiperazine ([001] projection) after a direct phase determination with texture electron-diffraction intensity data obtained originally by Vainshtein (1955).

can be particularly effective for structure analysis. Of particular importance historically have been the Σ_2-triple invariants where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$ and $\mathbf{h}_1 \neq \mathbf{h}_2 \neq \mathbf{h}_3$. The probability of predicting $\psi = 0$ is directly related to the value of

$$A = (2\sigma_3/\sigma_2^{3/2})|E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3}|,$$

where $\sigma_{\mathbf{h}} = \sum_{j=1}^{N} Z_j^n$ and Z is the value of the scattering factor at $\sin\theta/\lambda = 0$. Thus, the values of the phases are related to the measured structure factors, just as they are found to be in X-ray crystallography. The normalization described above imposes the point-atom structure (compensating for the fall-off of an approximately Gaussian form factor) often assumed in deriving the joint probability distributions. Especially for van der Waals structures, the constraint of positivity also holds in electron crystallography. (It is also quite useful for charged atoms so long as the reflections are not measured at very low angles.) Other useful phase invariant sums are the Σ_1 triples, where $\mathbf{h}_1 = \mathbf{h}_2 = -1/2\mathbf{h}_3$, and the quartets, where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$ and $\mathbf{h}_1 \neq \mathbf{h}_2 \neq \mathbf{h}_3 \neq \mathbf{h}_4$. The prediction of a correct phase for an invariant is related in each case to the normalized structure-factor magnitudes.

The procedure for phase determination, therefore, is identical to the one used in X-ray crystallography (see Chapter 2.2). Using vectorial combinations of Miller indices, one generates triple and quartet invariants from available measured data and ranks them according to parameters such as A, defined above, which, as shown in Chapter 2.2, are arguments of the Cochran formula. The invariants are thus listed in order of their reliability. This, in fact, generates a set of simultaneous equations in crystallographic phase. In order to begin solving these equations, it is permissible to define arbitrarily the phase values of a limited number of reflections (three for a three-dimensional primitive unit cell) for reflections with Miller-index parity $hkl \neq ggg$ and $\sum_i h_i k_i l_i \neq ggg$, where g is an even number. This defines the origin of a unit cell. For noncentrosymmetric unit cells, the condition for defining the origin, which depends on the space group, is somewhat more complicated and an enantiomorph-defining reflection must be added.

In the evaluation of phase-invariant sums above a certain probability threshold, phase values are determined algebraically after origin (and enantiomorph) definition until a large enough set is obtained to permit calculation of an interpretable potential map (*i.e.* where atomic positions can be seen). There may be a few invariant phase sums above this threshold probability value which are incorrectly predicted, leading either to false phase assignments or at least to phase assignments inconsistent with those found from other invariants. A small number of such errors can generally be tolerated. Another problem arises when an insufficient quantity of new phase values is assigned directly from the phase invariants after the origin-defining phases are defined. This difficulty may occur for small data sets, for example. If this is the case, it is possible that a new reflection of proper index parity can be used to define the origin. Alternatively, $\phi_n = a, b, c \ldots$ algebraic unknowns can be used to establish the phase linkage among certain reflections. If the structure is centrosymmetric, and when enough reflections are given at least symbolic phase assignments, 2^n maps are calculated and the correct structure is identified by inspection of the potential maps. When all goes well in this so-called 'symbolic addition' procedure, the symbols are uniquely determined and there is no need to calculate more than a single map. If algebraic values are retained for certain phases because of limited vectorial connections in the data set, then a few maps may need to be generated so that the correct structure can be identified using the chemical knowledge of the investigator. The atomic positions identified can then be used to calculate phases for all observed data (*via* the structure-factor calculation) and the structure can be refined by Fourier (or, sometimes, least-squares) techniques to minimize the crystallographic R factor.

The first actual application of direct phasing techniques to experimental electron-diffraction data, based on symbolic addition procedures, was to two methylene subcell structures (an *n*-paraffin and a phospholipid; Dorset & Hauptman, 1976). Since then, evaluation of phase invariants has led to numerous other structures. For example, early texture electron-diffraction data sets obtained in Moscow (Vainshtein, 1964) were shown to be suitable for direct analysis. The structure of diketopiperazine (Dorset, 1991*a*) was determined from these electron-diffraction data (Vainshtein, 1955) when directly determined phases allowed computation of potential maps such as the one shown in Fig. 2.5.8.1. Bond distances and angles are in good agreement with the X-ray structure, particularly after least-squares refinement (Dorset & McCourt, 1994*a*). In addition, the structures of urea (Dorset, 1991*b*), using data published by Lobachev & Vainshtein (1961), paraelectric thiourea (Dorset, 1991*b*), using data published by Dvoryankin & Vainshtein (1960), and three mineral structures (Dorset, 1992*a*), from data published by Zvyagin (1967), have been determined, all using the original texture (or mosaic single-crystal) diffraction data. The most recent determination based on such texture diffraction data is that of basic copper chloride (Voronova & Vainshtein, 1958; Dorset, 1994*c*).

Symbolic addition has also been used to assign phases to selected-area diffraction data. The crystal structure of boric acid (Cowley, 1953) has been redetermined, adding an independent low-temperature analysis (Dorset, 1992*b*). Additionally, a direct structure analysis has been reported for graphite, based on high-voltage intensity data (Ogawa *et al.*, 1994). Two-dimensional data from several polymer structures have also been analysed successfully (Dorset, 1992*c*) as have three-dimensional intensity data (Dorset, 1991*c*,*d*; Dorset & McCourt, 1993).

Phase information from electron micrographs has also been used to aid phase determination by symbolic addition. Examples include the epitaxically oriented paraffins *n*-hexatriacontane (Dorset & Zemlin, 1990), *n*-tritriacontane (Dorset & Zhang, 1991) and a 1:1 solid solution of n-$C_{32}H_{66}$/n-$C_{36}H_{74}$ (Dorset,

1990*a*). Similarly, lamellar electron-diffraction data to *ca* 3 Å resolution from epitaxically oriented phospholipids have been phased by analysis of Σ_1 and Σ_2-triplet invariants (Dorset, 1990*b*, 1991*e*,*f*), in one case combined with values from a 6 Å resolution electron-microscope image (Dorset *et al.*, 1990, 1993). Most recently, such data have been used to determine the layer packing of a phospholipid binary solid solution (Dorset, 1994*d*).

An *ab initio* direct phase analysis was carried out with zonal electron-diffraction data from copper perchlorophthalocyanine. Using intensities from a *ca* 100 Å thick sample collected at 1.2 MeV, the best map from a phase set with symbolic unknowns retrieves the positions of all the heavy atoms, equivalent to the results of the best images (Uyeda *et al.*, 1978–1979). Using these positions to calculate an initial phase set, the positions of the remaining light C, N atoms were found by Fourier refinement so that the final bond distances and angles were in good agreement with those from X-ray structures of similar compounds (Dorset *et al.*, 1991). A similar analysis has been carried out for the perbromo analogue (Dorset *et al.*, 1992). Although dynamical scattering and secondary scattering significantly perturb the observed intensity data, the total molecular structure can be visualized after a Fourier refinement. Most recently, a three-dimensional structure determination was reported for C_{60} buckminsterfullerene based on symbolic addition with results most in accord with a rotationally disordered molecular packing (Dorset & McCourt, 1994*b*).

2.5.8.4. The tangent formula

Given a triple phase relationship

$$\phi_{\mathbf{h}} \simeq \phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}},$$

where $\mathbf{h}$, $\mathbf{k}$ and $\mathbf{h}-\mathbf{k}$ form a vector sum, it is often possible to find a more reliable estimate of $\phi_{\mathbf{h}}$ when all the possible vectorial contributions to it within the observed data set $\mathbf{k}_{\mathbf{r}}$ are considered as an average, *viz:*

$$\phi_{\mathbf{h}} \simeq \langle \phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}} \rangle_{\mathbf{k}_{\mathbf{r}}}.$$

For actual phase determination, this can be formalized as follows. After calculating normalized structure-factor magnitudes $|E_{\mathbf{h}}|$ from the observed $|F_{\mathbf{h}}|$ to generate all possible phase triples within a reasonably high $A_{\mathbf{h}}$ threshold, new phase values can be estimated after origin definition by use of the tangent formula (Karle & Hauptman, 1956):

$$\tan \phi_{\mathbf{h}} = \frac{\sum_{\mathbf{k}_{\mathbf{r}}} W_{\mathbf{h}} |E_{\mathbf{k}}||E_{\mathbf{h}-\mathbf{k}}| \sin(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}_{\mathbf{r}}} W_{\mathbf{h}} |E_{\mathbf{k}}||E_{\mathbf{h}-\mathbf{k}}| \cos(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})}.$$

The reliability of the phase estimate depends on the variance $V(\phi_{\mathbf{h}})$, which is directly related to the magnitude of $\alpha_{\mathbf{h}}$, *i.e.*

$$\alpha_{\mathbf{h}}^2 = \left[\sum_{\mathbf{k}_{\mathbf{r}}} A_{\mathbf{h},\mathbf{k}} \cos(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})\right]^2 + \left[\sum_{\mathbf{k}_{\mathbf{r}}} A_{\mathbf{h},\mathbf{k}} \sin(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})\right]^2;$$

$A_{\mathbf{h},\mathbf{k}}$ is identical to the A value defined in the previous section. In the initial stages of phase determination $\alpha_{\mathbf{h}}$ is replaced by an expectation value α_{E} until enough phases are available to permit its calculation.

The phase solutions indicated by the tangent formula can thus be ranked according to the phase variance and the determination of phases can be made symbolically from the most probable triple-product relationships. This procedure is equivalent to the one described above for the evaluation of phase-invariant sums by symbolic addition. This procedure may allow determination of a large enough basis phase set to produce an interpretable map.

An alternative procedure is to use an automatic version of the tangent formula in a multisolution process. This procedure is described in Chapter 2.2. After origin definition, enough algebraic unknowns are defined (two values if centrosymmetric and four values, cycling through phase quadrants, if noncentrosymmetric) to access as many of the unknown phases as possible. These are used to generate a number of trial phase sets and the likelihood of identifying the correct solution is based on the use of some figure of merit.

Multisolution approaches employing the tangent formula include *MULTAN* (Germain *et al.*, 1971), *QTAN* (Langs & DeTitta, 1975) and *RANTAN* (Yao, 1981). *RANTAN* is a version of *MULTAN* that allows for a larger initial random phase set (with suitable control of weights in the tangent formula). *QTAN* utilizes the $\alpha_{\mathbf{h}_{\mathrm{est}}}$ definition, where

$$\alpha_{\mathbf{h}_{\mathrm{est}}} = \left\{\sum_{\mathbf{k}} A_{\mathbf{h},\mathbf{k}}^2 + 2 \sum_{\mathbf{k} \neq} \sum_{\mathbf{k}'} A_{\mathbf{h},\mathbf{k}} A'_{\mathbf{h},\mathbf{k}} \frac{I_1(A_{\mathbf{h},\mathbf{k}}) I_1(A'_{\mathbf{h},\mathbf{k}})}{I_0(A_{\mathbf{h},\mathbf{k}}) I_0(A'_{\mathbf{h},\mathbf{k}})}\right\}^{1/2},$$

for evaluating the phase variance. (Here I_0, I_1 are modified Bessel functions.) After multiple solutions are generated, it is desirable to locate the structurally most relevant phase sets by some figure of merit. There are many that have been suggested (Chapter 2.2). The most useful figure of merit in *QTAN* has been the NQEST (De Titta *et al.*, 1975) estimate of negative quartet invariants (see Chapter 2.2). More recently, this has been superseded by the minimal function (Hauptman, 1993):

$$R(\phi) = \frac{\sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{h},\mathbf{k}} (\cos \phi_{\mathbf{h},\mathbf{k}} - t_{\mathbf{h},\mathbf{k}})^2}{\sum_{\mathbf{h},\mathbf{k}} A_{\mathbf{h},\mathbf{k}}},$$

where $t_{\mathbf{h},\mathbf{k}} = I_1(A_{\mathbf{h},\mathbf{k}})/I_0(A_{\mathbf{h},\mathbf{k}})$ and $\phi_{\mathbf{h},\mathbf{k}} = \phi_{\mathbf{h}} + \phi_{\mathbf{k}} + \phi_{-\mathbf{h}-\mathbf{k}}$.

In the first application (Dorset *et al.*, 1979) of multisolution phasing to electron-diffraction data (using the program *QTAN*), *n*-beam dynamical structure factors generated for cytosine and disodium 4-oxypyrimidine-2-sulfinate were used to assess the effect of increasing crystal thickness and electron accelerating voltage on the success of the structure determination. At 100 kV samples at least 80 Å thick were usable for data collection and at 1000 kV this sample thickness limit could be pushed to 300 Å – or, perhaps, 610 Å if a partial structure were accepted for later Fourier refinement. NQEST identified the best phase solutions. Later *QTAN* was used to evaluate the effect of elastic crystal bend on the structure analysis of cytosine (Moss & Dorset, 1982).

In actual experimental applications, two forms of thiourea were investigated with *QTAN* (Dorset, 1992*d*), using published three-dimensional electron-diffraction intensities (Dvoryankin & Vainshtein, 1960, 1962). Analysis of the centrosymmetric paraelectric structure yielded results equivalent to those found earlier by symbolic addition (Dorset, 1991*b*). Analysis of the noncentrosymmetric ferroelectric form was also quite successful (Dorset, 1992*d*). In both cases, the correct structure was found at the lowest value of NQEST. Re-analysis of the diketopiperazine structure with *QTAN* also found the correct structure (Dorset & McCourt, 1994*a*) within the four lowest values of NQEST, but not the one at the lowest value. The effectiveness of this figure of merit became more questionable when *QTAN* was used to solve the noncentrosymmetric crystal structure of a polymer (Dorset, McCourt, Kopp *et al.*, 1994). The solution could not be found readily when NQEST was used but was easily identified when the minimal function $R(\phi)$ was employed instead.

MULTAN has been used to phase simulated data from copper perchlorophthalocyanine (Fan *et al.*, 1985). Within the 2 Å

resolution of the electron-microscope image, if one seeks phases for diffraction data in reciprocal-space regions where the objective lens phase contrast transfer function $|C(s)| \leq 0.2$, the method proves to be successful. The method is also quite effective for phase extension from 2 Å to 1 Å diffraction resolution, where the low-angle data serve as a large initial phase set for the tangent formula. However, no useful results were found from an *ab initio* phase determination carried out solely with the electron-diffraction structure-factor magnitudes. Similar results were obtained when *RANTAN* was used to phase experimental data from this compound (Fan *et al.*, 1991), *i.e.* the multisolution approach worked well for phase extension but not for *ab initio* phase determination. Additional tests were subsequently carried out with *QTAN* on an experimental *hk*0 electron-diffraction data set collected at 1200 kV (Dorset, McCourt, Fryer *et al.*, 1994). Again, *ab initio* phase determination is not possible by this technique. However, if a basis set was constructed from the Fourier transform of a 2.4 Å image, a correct solution could be found, but not at the lowest value of NQEST. This figure of merit was useful, however, when the basis set was taken from the symbolic addition determination mentioned in the previous section.

2.5.8.5. Density modification

Another method of phase determination, which is best suited to refining or extending a partial phase set, is the Hoppe–Gassmann density modification procedure (Hoppe & Gassmann, 1968; Gassmann & Zechmeister, 1972; Gassmann, 1976). The procedure is very simple but also very computer-intensive. Starting with a small set of (phased) $F_{\mathbf{h}}$, an initial potential map $\varphi(\mathbf{r})$ is calculated by Fourier transformation. This map is then modified by some real-space function, which restricts peak sizes to a maximum value and removes all negative density regions. The modified map $\varphi'(\mathbf{r})$ is then Fourier-transformed to produce a set of phased structure factors. Phase values are accepted *via* another modification function in reciprocal space, *e.g.* $E_{\rm calc}/E_{\rm obs} \geq p$, where p is a threshold quantity. The new set is then transformed to obtain a new $\varphi(\mathbf{r})$ and the phase refinement continues iteratively until the phase solution converges (judged by lower crystallographic R values).

The application of density modification procedures to electron-crystallographic problems was assessed by Ishizuka *et al.* (1982), who used simulated data from copper perchlorophthalocyanine within the resolution of the electron-microscope image. The method was useful for finding phase values in reciprocal-space regions where the transfer function $|C(s)| \leq 0.2$. As a technique for phase extension, density modification was acceptable for test cases where the resolution was extended from 1.67 to 1.0 Å, or 2.01 to 1.21 Å, but it was not very satisfactory for a resolution enhancement from 2.5 to 1.67 Å. There appear to have been no tests of this method yet with experimental data. However, the philosophy of this technique will be met again below in the description of the maximum entropy and likelihood procedure.

2.5.8.6. Convolution techniques

One of the first relationships ever derived for phase determination is the Sayre (1952) equation:

$$F_{\mathbf{h}} = \frac{\theta}{V} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}},$$

which is a simple convolution of phased structure factors multiplied by a function of the atomic scattering factors. For structures with nonoverlapping atoms, consisting of one atomic species, it is an exact expression. Although the convolution term resembles part of the tangent formula above, no statistical averaging is implied (Sayre, 1980). In X-ray crystallography this relationship has not been used very often, despite its accuracy. Part of the reason for this is that it requires relatively high resolution data for it to be useful. It can also fail for structures comprised of different atomic species.

Since, relative to X-ray scattering factors, electron scattering factors span a narrower range of magnitudes at $\sin\theta/\lambda = 0$, it might be thought that the Sayre equation would be particularly useful in electron crystallography. In fact, Liu *et al.* (1988) were able to extend phases for simulated data from copper perchlorophthalocyanine starting at the image resolution of 2 Å and reaching the 1 Å limit of an electron-diffraction data set. This analysis has been improved with a 2.4 Å basis set obtained from the Fourier transform of an electron micrograph of this material at 500 kV and extended to the 1.0 Å limit of a 1200 kV electron-diffraction pattern (Dorset *et al.*, 1995). Using the partial phase sets for zonal diffraction data from several polymers by symbolic addition (see above), the Sayre equation has been useful for extending into the whole *hk*0 set, often with great accuracy. The size of the basis set is critical but the connectivity to access all reflections is more so. Fan and co-workers have had considerable success with the analysis of incommensurately modulated structures. The average structure (basis set) is found by high-resolution electron microscopy and the 'superlattice' reflections, corresponding to the incommensurate modulation, are assigned phases in hyperspace by the Sayre convolution. Examples include a high T_c superconductor (Mo *et al.*, 1992) and the mineral ankangite (Xiang *et al.*, 1990). Phases of regular inorganic crystals have also been extended from the electron micrograph to the electron-diffraction resolution by this technique (Hu *et al.*, 1992).

In an investigation of how direct methods might be used for phase extension in protein electron crystallography, low-resolution phases from two proteins, bacteriorhodopsin (Henderson *et al.*, 1986) and halorhodopsin (Havelka *et al.*, 1993) were extended to higher resolution with the Sayre equation (Dorset *et al.*, 1995). For the noncentrosymmetric bacteriorhodopsin *hk*0 projection a 10 Å basis set was used, whereas a 15 Å set was accepted for the centrosymmetric halorhodopsin projection. In both cases, extensions to 6 Å resolution were reasonably successful. For bacteriorhodopsin, for which data were available to 3.5 Å, problems with the extension were encountered near 5 Å, corresponding to a minimum in a plot of average intensity *versus* resolution. Suggestions were made on how a multisolution procedure might be successful beyond this point.

2.5.8.7. Maximum entropy and likelihood

Maximum entropy has been applied to electron crystallography in several ways. In the sense that images are optimized, the entropy term

$$S = -\sum_i P_i \ln P_i,$$

where $P_i = p_i / \sum_i p_i$ and p_i is a pixel density, has been evaluated for various test electron-microscope images. For crystals, the true projected potential distribution function is thought to have the maximum value of S. If the phase contrast transfer function used to obtain a micrograph is unknown, test images (*i.e.* trial potential maps) can be calculated for different values of $\Delta f_{\rm trial}$. The value that corresponds to the maximum entropy would be near the true defocus. In this way, the actual objective lens transfer function can be found for a single image (Li, 1991) in addition to the other techniques suggested by this group.

Another use of the maximum-entropy concept is to guide the progress of a direct phase determination (Bricogne & Gilmore, 1990; Gilmore *et al.*, 1990). Suppose that there is a small set H of known phases $\phi_{\mathbf{h}\in H}$ (corresponding either to origin

definition, or the Fourier transform of an electron micrograph, or both) with associated unitary structure-factor amplitudes $|U_{\mathbf{h}\in H}|$. [The unitary structure factor is defined as $|U_{\mathbf{h}}| = |E_{\mathbf{h}}|/(N)^{1/2}$.] As usual, the task is to expand into the unknown phase set K to solve the crystal structure. From Bayes' theorem, the procedure is based on an operation where $p(\text{map}|\text{data}) \propto p(\text{map})p(\text{data}|\text{map})$. This means that the probability of successfully deriving a potential map, given diffraction data, is estimated. This so-called posterior probability is approximately proportional to the product of the probability of generating the map (known as the prior) and the probability of generating the data, given the map (known as the likelihood). The latter probability consults the observed data and can be used as a figure of merit.

Beginning with the basis set H, a trial map is generated from the limited number of phased structure factors. As discussed above, the map can be immediately improved by removing all negative density. The map can be improved further if its entropy is maximized using the equation given above for S. This produces the so-called maximum-entropy prior $q^{\text{ME}}(X)$.

So far, it has been assumed that all $|U_{\mathbf{h}\in K}| = 0$. If large reflections from the K set are now added and their phase values are permuted, then a number of new maps can be generated and their entropies can be maximized as before. This creates a phasing 'tree' with many possible solutions; individual branch points can have further reflections added *via* permutations to produce further sub-branches, and so on. Obviously, some figure of merit is needed to 'prune' the tree, *i.e.* to find likely paths to a solution.

The desired figure of merit is the likelihood $L(H)$. First a quantity

$$\Lambda_{\mathbf{h}} = 2NR \exp[-N(r^2 + R^2)] I_o(2NrR),$$

where $r = |^{\text{ME}}U_{\mathbf{h}}|$ (the calculated unitary structure factors) and $R = |^{o}U_{\mathbf{h}}|$ (the observed unitary structure factors), is defined. From this one can calculate

$$L(H) = \sum_{\mathbf{h}\notin H} \ln \Lambda_{\mathbf{h}}.$$

The null hypothesis $L(H_o)$ can also be calculated from the above when $r = 0$, so that the likelihood gain

$$LLg = L(H) - L(H_o)$$

ranks the nodes of the phasing tree in order of the best solutions.

Applications have been made to experimental electron-crystallographic data. A small-molecule structure starting with phases from an electron micrograph and extending to electron-diffraction resolution has been reported (Dong *et al.*, 1992). Other experimental electron-diffraction data sets used in other direct phasing approaches (see above) also have been assigned phases by this technique (Gilmore, Shankland & Bricogne, 1993). These include intensities from diketopiperazine and basic copper chloride. An application of this procedure in protein structure analysis has been published by Gilmore *et al.* (1992) and Gilmore, Shankland & Fryer (1993). Starting with 15 Å phases, it was possible to extend phases for bacteriorhodopsin to the limits of the electron-diffraction pattern, apparently with greater accuracy than possible with the Sayre equation (see above).

2.5.8.8. Influence of multiple scattering on direct electron-crystallographic structure analysis

The aim of electron-crystallographic data collection is to minimize the effect of dynamical scattering, so that the unit-cell potential distribution or its Fourier transform is represented significantly in the recorded signal. It would be a mistake, however, to presume that these data ever conform strictly to the kinematical approximation, for there is always some deviation from this ideal scattering condition that can affect the structure analysis. Despite this fact, some direct phasing procedures have been particularly 'robust', even when multiple scattering perturbations to the data are quite obvious (*e.g.* as evidenced by large crystallographic residuals).

The most effective direct phasing procedures seem to be those based on the Σ_2 triple invariants. These phase relationships will not only include the symbolic addition procedure, as it is normally carried out, but also the tangent formula and the Sayre equation (since it is well known that this convolution can be used to derive the functional form of the three-phase invariant). The strict ordering of $|E_{\mathbf{h}}|$ magnitudes is, therefore, not critically important so long as there are no major changes from large to small values (or *vice versa*). This was demonstrated in direct phase determinations of simulated n-beam dynamical diffraction data from a sulfur-containing polymer (Dorset & McCourt, 1992). Nevertheless, there is a point where measured data cannot be used. For example, intensities from *ca* 100 Å-thick epitaxically oriented copper perchlorophthalocyanine crystals become less and less representative of the unit-cell transform at lower electron-beam energies (Tivol *et al.*, 1993) and, accordingly, the success of the phase determination is compromised (Dorset, McCourt, Fryer *et al.*, 1994). The similarity between the Sayre convolution and the interactions of structure-factor terms in, *e.g.*, the multislice formulation of n-beam dynamical scattering was noted by Moodie (1965). It is interesting to note that dynamical scattering interactions observed by direct excitation of Σ_2 and Σ_1 triples in convergent-beam diffraction experiments can actually be exploited to determine crystallographic phases to very high precision (Spence & Zuo, 1992, pp. 56–63).

While the evaluation of positive quartet invariant sums (see Chapter 2.2) seems to be almost as favourable in the electron diffraction case as is the evaluation of Σ_2 triples, negative quartet invariants seem to be particularly sensitive to dynamical diffraction. If dynamical scattering can be modelled crudely by a convolutional smearing of the diffraction intensities, then the lowest structure-factor amplitudes, and hence the estimates of lowest $|E_{\mathbf{h}}|$ values, will be the ones most compromised. Since the negative quartet relationships require an accurate prediction of small 'cross-term' $|E_{\mathbf{h}}|$ values, multiple scattering can, therefore, limit the efficacy of this invariant for phase determination. In initial work, negative quartets have been mostly employed in the NQEST figure of merit, and analyses (Dorset, McCourt, Fryer *et al.*, 1994; Dorset & McCourt, 1994*a*) have shown how the degradation of weak kinematical $|E_{\mathbf{h}}|$ terms effectively reduced its effectiveness for locating correct structure solutions *via* the tangent formula, even though the tangent formula itself (based on triple phase estimates) was quite effective for phase determination. Substitution of the minimal function $R(\phi)$ for NQEST seems to have overcome this difficulty. [It should be pointed out, though, that only the Σ_2-triple contribution to $R(\phi)$ is considered.]

Structure refinement is another area where the effects of dynamical scattering are also problematic. For example, in the analysis of the paraelectric thiourea structure (Dorset, 1991*b*) from published texture diffraction data (Dvoryankin & Vainshtein, 1960), it was virtually impossible to find a chemically reasonable structure geometry by Fourier refinement, even though the direct phase determination itself was quite successful. The best structure was found only when higher-angle intensities (*i.e.* those least affected by dynamical scattering) were used to generate the potential map. Later analyses on heavy-atom-containing organics (Dorset *et al.*, 1992) found that the lowest kinematical R-factor value did not correspond to the chemically correct structure geometry. This observation was also made in the

least-squares refinement of diketopiperazine (Dorset & McCourt, 1994*a*). It is obvious that, if a global minimum is sought for the crystallographic residual, then dynamical structure factors, rather than kinematical values, should be compared to the observed values (Dorset *et al.*, 1992). Ways of integrating such calculations into the refinement process have been suggested (Sha *et al.*, 1993). Otherwise one must constrain the refinement to chemically reasonable bonding geometry in a search for a local *R*-factor minimum.

Corrections for such deviations from the kinematical approximation are complicated by the presence of other possible data perturbations, especially if microareas are being sampled, *e.g.* in typical selected-area diffraction experiments. Significant complications can arise from the diffraction incoherence observed from elastically deformed crystals (Cowley, 1961) as well as secondary scattering (Cowley *et al.*, 1951). These complications were also considered for the larger (*e.g.* millimetre diameter) areas sampled in an electron-diffraction camera when recording texture diffraction patterns (Turner & Cowley, 1969), but, because of the crystallite distributions, it is sometimes found that the two-beam dynamical approximation is useful (accounting for a number of successful structure analyses carried out in Moscow).

MT is grateful to Professor J. C. H. Spence for critical reading of the manuscript of Section 2.5.3, for correcting the English and for helpful comments, and to Professor M. Terauchi and Dr K. Tsuda of Tohoku University and Dr K. Saitoh of Nagoya University for useful discussions.

References

Adiga, P. S., Malladi, R., Baxter, W. & Glaeser, R. M. (2004). *A binary segmentation approach for boxing ribosome particles in cryo EM micrographs*. *J. Struct. Biol.* **145**, 142–151.

Agrawal, R. K., Penczek, P., Grassucci, R. A., Li, Y., Leith, A., Nierhaus, K. H. & Frank, J. (1996). *Direct visualization of A-, P-, and E-site transfer RNAs in the Escherichia coli ribosome*. *Science*, **271**, 1000–1002.

Avilov, A. S. (1979). *Electrical measurement of reflection intensities on electron diffraction from mosaic single crystals. Sov. Phys. Crystallogr.* **24**, 103–104.

Avilov, A. S., Kuligin, A. K., Pietsch, U., Spence, J. C. H., Tsirelson, V. G. & Zuo, J. M. (1999). *Scanning system for high-energy electron diffractometry. J. Appl. Cryst.* **32**, 1033–1038.

Avilov, A. S., Parmon, V. S., Semiletov, S. A. & Sirota, M. I. (1984). *Calculation of reflected intensities in multiple-beam diffraction of fast electrons by polycrystalline specimens. Sov. Phys. Crystallogr.* **29**, 5–7.

Baldwin, P. R. & Penczek, P. A. (2007). *The transform class in SPARX and EMAN2*. *J. Struct. Biol.* **157**, 250–261.

Basokur, A. T. (1998). *Digital filter design using the hyperbolic tangent functions*. *J. Balkan Geophys. Soc.* **1**, 14–18.

Bendersky, L. A. (1985). *Quasicrystal with one-dimensional translational symmetry and a tenfold rotation axis*. *Phys. Rev. Lett.* **55**, 1461–1467.

Bendersky, L. A. (1986). *Decagonal phase. J. Phys. Colloq.* **47**, C3, 457–464.

Bendersky, L. A. & Kaufman, M. J. (1986). *Determination of the point group of the icosahedral phase in an Al–Mn–Si alloy using convergent-beam electron diffraction*. *Philos. Mag.* **B53**, L75–L80.

Bethe, H. A. (1928). *Theorie der Beugung von Elektronen an Kristallen. Ann. Phys. (Leipzig)*, **87**, 55–129.

Biemond, J., Lagendijk, R. L. & Mersereau, R. M. (1990). *Iterative methods for image deblurring*. *Proc. IEEE*, **78**, 856–883.

Blackman, M. (1939). *On the intensities of electron diffraction rings. Proc. R. Soc. London Ser. A*, **173**, 68–82.

Boisset, N., Penczek, P., Pochon, F., Frank, J. & Lamy, J. (1993). *Three-dimensional architecture of human alpha 2-macroglobulin transformed with methylamine*. *J. Mol. Biol.* **232**, 522–529.

Boisset, N., Penczek, P., Taveau, J. C., Lamy, J. & Frank, J. (1995). *Three-dimensional reconstruction of Androctonus australis hemocyanin labeled with a monoclonal Fab fragment. J. Struct. Biol.* **115**, 16–29.

Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy*. *Nature (London)*, **386**, 88–91.

Bracewell, R. N. (1956). *Strip integration in radio astronomy. Austr. J. Phys*. **9**, 198–217.

Bracewell, R. N. & Riddle, A. C. (1967) *Inversion of fan-beam scans in radio astronomy. Astrophys. J.* **150**, 427–434.

Bricogne, G. & Gilmore, C. J. (1990). *A multisolution method of phase determination by combined maximization of entropy and likelihood. I. Theory, algorithms and strategy. Acta Cryst*. A**46**, 284–297.

Brisse, F. (1989). *Electron diffraction of synthetic polymers: the model compound approach to polymer structure. J. Electron Microsc. Tech*. **11**, 272–279.

Buxton, B., Eades, J. A., Steeds, J. W. & Rackham, G. M. (1976). *The symmetry of electron diffraction zone axis patterns. Philos. Trans. R. Soc. London Ser. A*, **181**, 171–193.

Carazo, J. M. (1992). *The fidelity of 3D reconstruction from incomplete data and the use of restoration methods*. In *Electron Tomography*, edited by J. Frank, pp. 117–166. New York: Plenum.

Carazo, J. M. & Carrascosa, J. L. (1986). *Information recovery in missing angular data cases: an approach by the convex projections method in three dimensions*. *J. Microsc.* **145**, 23–43.

Carazo, J. M. & Carrascosa, J. L. (1987). *Restoration of direct Fourier three-dimensional reconstructions of crystalline specimens by the method of convex projections*. *J. Microsc.* **145**, 159–177.

Cherns, D., Kiely, C. J. & Preston, A. R. (1988). *Electron diffraction studies of strain in epitaxial bicrystals and multilayers*. *Ultramicroscopy*, **24**, 355–370.

Cherns, D. & Preston, A. R. (1986). *Convergent beam diffraction studies of crystal defects*. *Proc. XI Int. Congr. Electron Microsc.*, Kyoto, Japan, p. 721.

Clark, J. J., Palmer, M. R. & Lawrence, P. D. (1985). *A transformation method for the reconstruction of functions from nonuniformly spaced samples*. *IEEE Trans. Acoust. Speech Signal Process*. **33**, 1151–1165.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *The structure of synthetic polypeptides. 1. The transform of atoms on a helix. Acta Cryst*. **5**, 581–586.

Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). *Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy*. *Nature (London)*, **386**, 91–94.

Cowley, J. M. (1953). *Structure analysis of single crystals by electron diffraction. II. Disordered boric acid structure. Acta Cryst*. **6**, 522–529.

Cowley, J. M. (1956). *A modified Patterson function. Acta Cryst*. **9**, 397–398.

Cowley, J. M. (1961). *Diffraction intensities from bent crystals. Acta Cryst*. **14**, 920–927.

Cowley, J. M. (1981). *Diffraction Physics*, 2nd ed. Amsterdam: North-Holland.

Cowley, J. M. (1995). *Diffraction Physics*, 3rd ed. Amsterdam: North-Holland.

Cowley, J. M. & Au, A. Y. (1978). *Image signals and detector configurations for STEM*. In *Scanning Electron Microscopy*, Vol. 1, pp. 53–60. AFM O'Hare: SEM Inc.

Cowley, J. M. & Moodie, A. F. (1957). *The scattering of electrons by atoms and crystals. I. A new theoretical approach. Acta Cryst*. **10**, 609–619.

Cowley, J. M. & Moodie, A. F. (1959). *The scattering of electrons by atoms and crystals. III. Single-crystal diffraction patterns. Acta Cryst*. **12**, 360–367.

Cowley, J. M. & Moodie, A. F. (1960). *Fourier images. IV. The phase grating. Proc. Phys. Soc. London*, **76**, 378–384.

Cowley, J. M., Moodie, A. F., Miyake, S., Takagi, S. & Fujimoto, F. (1961). *The extinction rules for reflections in symmetrical electron diffraction spot patterns. Acta Cryst*. **14**, 87–88.

Cowley, J. M., Rees, A. L. G. & Spink, J. A. (1951). *Secondary elastic scattering in electron diffraction. Proc. Phys. Soc. London Sect. A*, **64**, 609–619.

Cramér, H. (1954). *Mathematical Methods of Statistics*. Princeton University Press.

Crowther, R. A. & Amos, L. A. (1971). *Harmonic analysis of electron microscope images with rotational symmetry. J. Mol. Biol*. **60**, 123–130.

Crowther, R. A. & Amos, L. A. (1972). *Three-dimensional image reconstructions of some small spherical viruses*. *Cold Spring Harbor Symp. Quant. Biol.* **36**, 489–494.

Crowther, R. A., Amos, L. A., Finch, J. T., DeRosier, D. J. & Klug, A. (1970). *Three dimensional reconstruction of spherical viruses by*

Fourier synthesis from electron micrographs. Nature (London), **226**, 421–425.

Crowther, R. A., DeRosier, D. J. & Klug, A. (1970). *The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc. R. Soc. London Ser. A*, **317**, 319–340.

Dawson, B., Goodman, P., Johnson, A. W. S., Lynch, D. F. & Moodie, A. F. (1974). *Some definitions and units in electron diffraction. Acta Cryst*. A**30**, 297–298.

De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Use of negative quartet cosine invariants as a phasing figure of merit: NQEST. Acta Cryst*. A**31**, 472–479.

DeRosier, D. J. & Klug, A. (1968). *Reconstruction of three dimensional structures from electron micrographs. Nature (London)*, **217**, 130–134.

Dong, W., Baird, T., Fryer, J. R., Gilmore, C. J., MacNicol, D. D., Bricogne, G., Smith, D. J., O'Keefe, M. A. & Hovmöller, S. (1992). *Electron microscopy at 1 Å resolution by entropy maximization and likelihood ranking. Nature (London)*, **355**, 605–609.

Dorset, D. L. (1976). *The interpretation of quasi-kinematical single-crystal electron diffraction intensity data from paraffins. Acta Cryst*. A**32**, 207–215.

Dorset, D. L. (1987). *Electron diffraction structure analysis of phospholipids. J. Electron Microsc. Tech*. **7**, 35–46.

Dorset, D. L. (1990*a*). *Direct structure analysis of a paraffin solid solution. Proc. Natl Acad. Sci. USA*, **87**, 8541–8544.

Dorset, D. L. (1990*b*). *Direct determination of crystallographic phases for diffraction data from phospholipid multilamellar arrays. Biophys. J*. **58**, 1077–1087.

Dorset, D. L. (1991*a*). *Electron diffraction structure analysis of diketopiperazine – a direct phase determination. Acta Cryst*. A**47**, 510–515.

Dorset, D. L. (1991*b*). *Is electron crystallography possible? The direct determination of organic crystal structures. Ultramicroscopy*, **38**, 23–40.

Dorset, D. L. (1991*c*). *Electron diffraction structure analysis of polyethylene. A direct phase determination. Macromolecules*, **24**, 1175–1178.

Dorset, D. L. (1991*d*). *Electron crystallography of linear polymers: direct structure analysis of poly(ε-caprolactone). Proc. Natl Acad. Sci. USA*, **88**, 5499–5502.

Dorset, D. L. (1991*e*). *Direct determination of crystallographic phases for diffraction data from lipid bilayers. I. Reliability and phase refinement. Biophys. J*. **60**, 1356–1365.

Dorset, D. L. (1991*f*). *Direct determination of crystallographic phases for diffraction data from lipid bilayers. II. Refinement of phospholipid structures. Biophys. J*. **60**, 1366–1373.

Dorset, D. L. (1992*a*). *Direct phasing in electron crystallography: determination of layer silicate structures. Ultramicroscopy*, **45**, 5–14.

Dorset, D. L. (1992*b*). *Direct methods in electron crystallography – structure analysis of boric acid. Acta Cryst*. A**48**, 568–574.

Dorset, D. L. (1992*c*). *Electron crystallography of linear polymers: direct phase determination for zonal data sets. Macromolecules*, **25**, 4425–4430.

Dorset, D. L. (1992*d*). *Automated phase determination in electron crystallography: thermotropic phases of thiourea. Ultramicroscopy*, **45**, 357–364.

Dorset, D. L. (1994*a*). *Electron crystallography of organic molecules. Adv. Electron. Electron Phys*. **88**, 111–197.

Dorset, D. L. (1994*b*). *Electron crystallography of linear polymers*. In *Characterization of Solid Polymers. New Techniques and Developments*, edited by S. J. Spells, pp. 1–16. London: Chapman and Hall.

Dorset, D. L. (1994*c*). *Electron crystallography of inorganic compounds. Direct determination of the basic copper chloride structure $CuCl_2{\cdot}3Cu(OH)_2$. J. Chem. Crystallogr*. **24**, 219–224.

Dorset, D. L. (1994*d*). *Direct determination of layer packing for a phospholipid solid solution at 0.32 nm resolution. Proc. Natl Acad. Sci. USA*, **91**, 4920–4924.

Dorset, D. L., Beckmann, E. & Zemlin, F. (1990). *Direct determination of a phospholipid lamellar structure at 0.34 nm resolution. Proc. Natl Acad. Sci. USA*, **87**, 7570–7573.

Dorset, D. L. & Hauptman, H. A. (1976). *Direct phase determination for quasi-kinematical electron diffraction intensity data from organic microcrystals. Ultramicroscopy*, **1**, 195–201.

Dorset, D. L., Jap, B. K., Ho, M.-H. & Glaeser, R. M. (1979). *Direct phasing of electron diffraction data from organic crystals: the effect of n-beam dynamical scattering. Acta Cryst*. A**35**, 1001–1009.

Dorset, D. L., Kopp, S., Fryer, J. R. & Tivol, W. F. (1995). *The Sayre equation in electron crystallography. Ultramicroscopy*, **57**, 59–89.

Dorset, D. L. & McCourt, M. P. (1992). *Effect of dynamical scattering on successful direct phase determination in electron crystallography – a model study. Trans. Am. Crystallogr. Assoc*. **28**, 105–113.

Dorset, D. L. & McCourt, M. P. (1993). *Electron crystallographic analysis of a polysaccharide structure – direct phase determination and model refinement for mannan I. J. Struct. Biol*. **111**, 118–124.

Dorset, D. L. & McCourt, M. P. (1994*a*). *Automated structure analysis in electron crystallography: phase determination with the tangent formula and least-squares refinement. Acta Cryst*. A**50**, 287–292.

Dorset, D. L. & McCourt, M. P. (1994*b*). *Disorder and molecular packing of C_{60} buckminsterfullerene: a direct electron-crystallographic analysis. Acta Cryst*. A**50**, 344–351.

Dorset, D. L., McCourt, M. P., Fryer, J. R., Tivol, W. F. & Turner, J. N. (1994). *The tangent formula in electron crystallography: phase determination of copper perchlorophthalocyanine. Microsc. Soc. Am. Bull*. **24**, 398–404.

Dorset, D. L., McCourt, M. P., Kopp, S., Wittmann, J.-C. & Lotz, B. (1994). *Direct determination of polymer crystal structures by electron crystallography – isotactic poly(1-butene), form III. Acta Cryst*. B**50**, 201–208.

Dorset, D. L., McCourt, M. P., Tivol, W. F. & Turner, J. N. (1993). *Electron diffraction from phospholipids – an approximate correction for dynamical scattering and tests for a correct phase determination. J. Appl. Cryst*. **26**, 778–786.

Dorset, D. L., Tivol, W. F. & Turner, J. N. (1991). *Electron crystallography at atomic resolution: ab initio structure analysis of copper perchlorophthalocyanine. Ultramicroscopy*, **38**, 41–45.

Dorset, D. L., Tivol, W. F. & Turner, J. N. (1992). *Dynamical scattering and electron crystallography – ab initio structure analysis of copper perbromophthalocyanine. Acta Cryst*. A**48**, 562–568.

Dorset, D. L. & Zemlin, F. (1990). *Direct phase determination in electron crystallography: the crystal structure of an n-paraffin. Ultramicroscopy*, **33**, 227–236.

Dorset, D. L. & Zhang, W. P. (1991). *Electron crystallography at atomic resolution: the structure of the odd-chain paraffin n-tritriacontane. J. Electron Microsc. Tech*. **18**, 142–147.

Dowell, W. C. T. & Goodman, P. (1973). *Image formation and contrast from convergent electron beam. Philos. Mag*. **28**, 471–473.

Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). *The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. EMBO J*. **12**, 1303–1309.

Dvoryankin, V. F. & Vainshtein, B. K. (1960). *An electron diffraction study of thiourea. Sov. Phys. Crystallogr*. **5**, 564–574.

Dvoryankin, V. F. & Vainshtein, B. K. (1962). *An electron diffraction study of the low-temperature ferroelectric form of thiourea. Sov. Phys. Crystallogr*. **6**, 765–772.

Eades, J. A. (1988). *Microbeam Analysis*, edited by D. Newbuly, p. 75. San Francisco Press.

Eades, J. A. & Spence, J. C. H. (1987). Private communication.

Egelman, E. (1986). *An algorithm for straightening images of curved filamentous structures. Ultramicroscopy*, **19**, 367–374.

Egelman, E. H. (2000). *A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy*, **85**, 225–234.

Egelman, E. H., Francis, N. & DeRosier, D. J. (1982). *F-actin is a helix with a random variable twist. Nature (London)*, **298**, 131–135.

Egelman, E. H. & Stasiak, A. (1988). *Structure of helical RecA–DNA complexes. II. Local conformational changes visualized in bundles of RecA–ATP gamma S filaments. J. Mol. Biol*. **200**, 329–349.

Erickson, H. P. & Klug, A. (1971). *Measurements and compensation of defocusing and aberrations by Fourier processing of electron micrographs. Philos. Trans. R. Soc. London Ser. B*, **261**, 105–118.

Fan, H. F., Xiang, S. B., Li, F. H., Pan, Q., Uyeda, N. & Fujiyoshi, Y. (1991). *Image resolution enhancement by combining information from electron diffraction pattern and micrograph. Ultramicroscopy*, **36**, 361–365.

Fan, H.-F., Zhong, Z.-Y., Zheng, C.-D. & Li, F.-H. (1985). *Image processing in high-resolution electron microscopy using the direct method. I. Phase extension. Acta Cryst*. A**41**, 163–165.

Farrow, N. A. & Ottensmeyer, F. P. (1992). *A posteriori determination of relative projection directions of arbitrarily oriented macromolecules. J. Opt. Soc. Am. A*, **9**, 1749–1760.

Fernandez, J.-J., Sanjurjo, J. R. & Carazo, J. M. (1997). *A spectral estimation approach to contrast transfer function detection in electron microscopy*. *Ultramicroscopy*, **68**, 267–295.

Fishman, G. (1995). *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer.

Frangakis, A. S., Bohm, J., Forster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R. & Baumeister, W. (2002). *Identification of macromolecular complexes in cryoelectron tomograms of phantom cells*. *Proc. Natl Acad. Sci. USA*, **99**, 14153–14158.

Frank, J. (1975). *Averaging of low exposure electron micrographs of non-periodic objects. Ultramicroscopy*, **1**, 159–162.

Frank, J. (1980). *The role of correlation techniques in computer image processing*. In *Computer Processing of Electron Microscopy Images*, edited by P. W. Hawkes, pp. 187–222. Berlin: Springer-Verlag.

Frank, J. (1990). *Classification of macromolecular assemblies studied as 'single particles'*. *Quart. Rev. Biophys.* **23**, 281–329.

Frank, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. New York: Oxford University Press.

Frank, J. & Mcewen, B. (1992). *Alignment by crosscorrelation*. In *Electron Tomography*, edited by J. Frank, pp. 205–214. New York: Plenum.

Frank, J., Penczek, P., Agrawal, R. K., Grassucci, R. A. & Heagle, A. B. (2000). *Three-dimensional cryoelectron microscopy of ribosomes*. *Methods Enzymol.* **317**, 276–291.

Frank, J., Penczek, P. & Liu, W. (1992). *Alignment, classification, and three-dimensional reconstruction of single particles embedded in ice*. *Scan. Microsc. Suppl.* **6**, 11–20.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). *SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields*. *J. Struct. Biol.* **116**, 190–199.

Frank, J., Radermacher, M., Wagenknecht, T. & Verschoor, A. (1988). *Studying ribosome structure by electron microscopy and computer-image processing*. *Methods Enzymol.* **164**, 3–35.

Frank, J., Verschoor, A. & Boublik, M. (1981). *Computer averaging of electron micrographs of 40S ribosomal subunits*. *Science*, **214**, 1353–1355.

Frank, J., Verschoor, A. & Boublik, M. (1982). *Multivariate statistical analysis of ribosome electron micrographs. L and R lateral views of the 40 S subunit from HeLa cells*. *J. Mol. Biol.* **161**, 107–133.

Frank, J. & Wagenknecht, T. (1984). *Automatic selection of molecular images from electron micrographs*. *Ultramicroscopy*, **12**, 169–176.

Fryer, J. R. (1993). *Electron crystallography of small organic molecules. Microsc. Soc. Am. Bull.* **23**, 44–56.

Fujimoto, F. (1959). *Dynamical theory of electron diffraction in Laue-case. I. General theory. J. Phys. Soc. Jpn*, **14**, 1158–1168.

Fujiwara, K. (1961). *Relativistic dynamical theory of electron diffraction. J. Phys. Soc. Jpn*, **16**, 2226–2238.

Fukuhara, A. (1966). *Many-ray approximations in the dynamical theory of electron diffraction. J. Phys. Soc. Jpn*, **21**, 2645–2662.

Fuller, S. D., Butcher, S. J., Cheng, R. H. & Baker, T. S. (1996). *Three-dimensional reconstruction of icosahedral particles – the uncommon line*. *J. Struct. Biol.* **116**, 48–55.

Fung, K. K., Steeds, J. W. & Eades, J. A. (1980). *Application of convergent beam electron diffraction to study the stacking of layers in transition-metal dichalcogenides*. *Physica (Utrecht)*, **B99**, 47–50.

Gabashvili, I. S., Agrawal, R. K., Spahn, C. M., Grassucci, R. A., Svergun, D. I., Frank, J. & Penczek, P. (2000). *Solution structure of the E. coli 70S ribosome at 11.5 Å resolution*. *Cell*, **100**, 537–549.

Gabor, D. (1949). *Microscopy by reconstructed wavefronts. Proc. R. Soc. London Ser. A*, **197**, 454–487.

Gassmann, J. (1976). *Improvement and extension of approximate phase sets in structure determination*. In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 144–154. Copenhagen: Munksgaard.

Gassmann, J. & Zechmeister, K. (1972). *Limits of phase expansion in direct methods. Acta Cryst.* A**28**, 270–280.

Germain, G., Main, P. & Woolfson, M. M. (1971). *The application of phase relationships to complex structures. III. The optimum use of phase relationships. Acta Cryst.* A**27**, 368–376.

Gilbert, P. F. C. (1972). *Iterative methods for the three-dimensional reconstruction of an object from projections. J. Theor. Biol.* **36**, 105–117.

Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *A multisolution method of phase determination by combined maximization of entropy and likelihood. II. Application to small molecules. Acta Cryst.* A**46**, 297–308.

Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Applications of the maximum entropy method to powder diffraction and electron crystallography. Proc. R. Soc. London Ser. A*, **442**, 97–111.

Gilmore, C. J., Shankland, K. & Fryer, J. R. (1992). *The application of the maximum entropy method to electron microscopy data for purple membrane. Trans. Am. Crystallogr. Assoc.* **28**, 129–139.

Gilmore, C. J., Shankland, K. & Fryer, J. R. (1993). *Phase extension in electron crystallography using the maximum entropy method and its application to two-dimensional purple membrane data from Halobacterium halobium. Ultramicroscopy*, **49**, 132–146.

Gjønnes, J. & Gjønnes, K. (1985). *Bloch wave symmetries and inclined surfaces*. *Ultramicroscopy*, **18**, 77–82.

Gjønnes, J. & Høier, R. (1971). *The application of non-systematic many-beam dynamic effects to structure-factor determination. Acta Cryst.* A**27**, 313–316.

Gjønnes, J. & Moodie, A. F. (1965). *Extinction conditions in dynamic theory of electron diffraction patterns. Acta Cryst.* **19**, 65–67.

Glauber, R. & Schomaker, V. (1953). *The theory of electron diffraction. Phys. Rev.* **89**, 667–670.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*. Baltimore: Johns Hopkins University Press.

Goncharov, A. B., Vainshtein, B. K., Ryskin, A. I. & Vagin, A. A. (1987). *Three-dimensional reconstruction of arbitrarily oriented identical particles from their electron photomicrographs. Sov. Phys. Crystallogr.* **32**, 504–509.

Gonzalez, R. F. & Woods, R. E. (2002). *Digital Image Processing*. Upper Saddle River: Prentice Hall.

Goodman, P. (1975). *A practical method of three-dimensional space-group analysis using convergent-beam electron diffraction*. *Acta Cryst.* A**31**, 804–810.

Goodman, P. & Johnson, A. W. S. (1977). *Identification of enantiomorphically related space groups by electron diffraction – a second method*. *Acta Cryst.* A**33**, 997–1001.

Goodman, P. & Lehmpfuhl, G. (1964). *Verbotene elektronenbeugungs-reflexe von CdS*. *Z. Naturforsch. Teil A*, **19**, 818–820.

Goodman, P. & Lehmpfuhl, G. (1965). *Elektronenbeugungsuntersuchungen im konvergenten bundel mit dem Siemens Elmiskop I. Z. Naturforsch. Teil A*, **20**, 110–114.

Goodman, P. & Lehmpfuhl, G. (1967). *Electron diffraction study of MgO h00-systematic interactions*. *Acta Cryst.* **22**, 14–24.

Goodman, P. & Secomb, T. W. (1977). *Identification of enantiomorphously related space groups by electron diffraction*. *Acta Cryst.* A**33**, 126–133.

Gordon, R., Bender, R. & Herman, G. T. (1970). *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography*. *J. Theor. Biol.* **29**, 471–481.

Grigorieff, N. (1998). *Three-dimensional structure of bovine NADH: ubiquinone oxidoreductase (complex I) at 22 Å in ice*. *J. Mol. Biol.* **277**, 1033–1046.

Grigorieff, N. (2007). *FREALIGN: High-resolution refinement of single particle structures*. *J. Struct. Biol.* **157**, 117.

Gurskaya, G. V., Lobanova, G. M. & Vainshtein, B. K. (1971). *X-ray diffraction and electron-microscope study of hexagonal catalase crystal. Sov. Phys. Crystallogr.* **16**, 662–669.

Hall, R. J. & Patwardhan, A. (2004). *A two step approach for semi-automated particle selection from low contrast cryo-electron micrographs*. *J. Struct. Biol.* **145**, 19–28.

Hamaker, C. & Solmon, D. C. (1978). *Angles between null spaces of X-rays*. *J. Math. Anal. Appl.* **62**, 1–23.

Harauz, G. & van Heel, M. (1986). *Exact filters for general geometry three-dimensional reconstruction*. *Optik*, **73**, 146–156.

Hashimoto, H., Endoh, H., Tanji, T., Ono, A. & Watanabe, E. (1977). *Direct observation of fine structure within images of atoms in crystals by transmission electron microscopy. J. Phys. Soc. Jpn*, **42**, 1073–1074.

Hashimoto, H., Mannami, M. & Naiki, T. (1961). *Dynamical theory of electron diffraction for the electron microscope image of crystal lattices. I. Image of single crystals. II. Image of superposed crystals (moiré pattern). Philos. Trans. R. Soc. London*, **253**, 459–516.

Hauptman, H. (1972). *Crystal Structure Determination. The Role of the Cosine Seminvariants*. New York: Plenum Press.

Hauptman, H. (1993). *A minimal principle in X-ray crystallography: starting in a small way. Proc. R. Soc. London Ser. A*, **442**, 3–12.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*. American Crystallographic Association Monograph No. 3. Ann Arbor: Edwards Brothers.

Havelka, W., Henderson, R., Heymann, J. A. W. & Oesterhelt, D. (1993). *Projection structure of halorhodopsin from Halobacterium halobium at 6 Å resolution obtained by electron cryomicroscopy. J. Mol. Biol*. **234**, 837–846.

He, L. X., Wu, Y. K. & Kuo, K. H. (1988). *Decagonal quasicrystals with different periodicities along the tenfold axis in rapidly solidified $Al_{65}Cu_{20}Mn_{15}$, $Al_{65}Cu_{20}Fe_{15}$, $Al_{65}Cu_{20}Co_{15}$ or $Al_{65}Cu_{20}Ni_{15}$. J. Mater. Sci. Lett.* **7**, 1284–1286.

Heel, M. van (1982). *Detection of objects in quantum-noise limited images*. *Ultramicroscopy*, **8**, 331–342.

Heel, M. van (1987*a*). *Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction*. *Ultramicroscopy*, **21**, 111–124.

Heel, M. van (1987*b*). *Similarity measures between images*. *Ultramicroscopy*, **21**, 95–100.

Heel, M. van, Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M. & Patwardhan, A. (2000). *Single-particle electron cryo-microscopy: towards atomic resolution*. *Quart. Rev. Biophys.* **33**, 307–369.

Heel, M. van, Harauz, G. & Orlova, E. V. (1996). *A new generation of the IMAGIC image processing system*. *J. Struct. Biol.* **116**, 17–24.

Heel, M. van & Hollenberg, J. (1980). *The stretching of distorted images of two-dimensional crystals*. In *Electron Microscopy at Molecular Dimensions*, edited by W. Baumeister, pp. 256–260. Berlin: Springer.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). *Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy*. *J. Mol. Biol*. **213**, 899–929.

Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. (1986). *Structure of purple membrane from Halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. Ultramicroscopy*, **19**, 147–178.

Henderson, R. & Unwin, P. N. T. (1975). *Three-dimensional model of purple membrane obtained by electron microscopy. Nature (London)*, **257**, 28–32.

Herman, G. T. & Meyer, L. B. (1993). *Algebraic reconstruction techniques can be made computationally efficient*. *IEEE Trans. Med. Imaging*, **12**, 600–609.

Herrmann, K. H., Krahl, D. & Rust, H.-P. (1980). *Low-dose image recording by TV techniques*. In *Electron Microscopy at Molecular Dimensions*, edited by W. Baumeister & W. Vogell, pp. 186–193. Berlin: Springer-Verlag.

Hirsch, P. B., Howie, A., Nicholson, R. B., Pashley, D. W. & Whelan, M. J. (1965). *Electron Microscopy of Thin Crystals*. London: Butterworths.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P. R., Huang, Z., Penczek, P. A., Yang, C., Glaeser, R. M., Adams, P. D. & Ludtke, S. J. (2007). *SPARX, a new environment for cryo-EM image processing*. *J. Struct. Biol.* **157**, 47–55.

Holmes, K. C., Angert, I., Kull, F. J., Jahn, W. & Schroder, R. R. (2003). *Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide*. *Nature (London)*, **425**, 423–427.

Hoppe, W. (1971). *Use of zone correction plate and other techniques for structure determination of aperiodic objects at atomic resolution using a conventional electron microscope. Philos. Trans. R. Soc. London Ser. B*, **261**, 71–94.

Hoppe, W., Bussler, P., Feltynowski, A., Hunsmann, N. & Hirt, A. (1973). *Some experience with computerized image reconstruction methods*. In *Image Processing and Computer-Aided Design in Electron Optics*, edited by R. W. Hawkes, pp. 92–126. London: Academic Press.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst*. B**24**, 97–107.

Horstmann, M. & Meyer, G. (1965). *Messung der Elektronenbeugungs-intensitaten polykristalliner Aluminium schichten bei tiefer Temperatur und Vergleich mit der dynamischen Theorie. Z. Phys*. **182**, 380–397.

Hovmöller, S., Sjögren, A., Farrants, G., Sundberg, M. & Marinder, B. O. (1984). *Accurate atomic positions from electron microscopy. Nature (London)*, **311**, 238–241.

Howe, J. M., Sarikaya, M. & Gronsky, R. (1986). *Space-group analyses of thin precipitates by different convergent-beam electron diffraction procedures*. *Acta Cryst.* A**42**, 368–380.

Hu, H. H., Li, F. H. & Fan, H. F. (1992). *Crystal structure determination of $K_2O{\cdot}7Nb_2O_5$ by combining high resolution electron microscopy and electron diffraction. Ultramicroscopy*, **41**, 387–397.

Huang, Z., Baldwin, P. R., Mullapudi, S. R. & Penczek, P. A. (2003). *Automated determination of parameters describing power spectra of micrograph images in electron microscopy*. *J. Struct. Biol.* **144**, 79–94.

Huang, Z. & Penczek, P. A. (2004). *Application of template matching technique to particle detection in electron micrographs*. *J. Struct. Biol.* **145**, 29–40.

Huesman, R. H., Gullberg, G. T., Greenberg, W. L. & Budinger, T. F. (1977). *RECLBL library users' manual – Donner algorithms for reconstruction tomography*. University of California, Berkeley, USA.

Hurley, A. C. & Moodie, A. F. (1980). *The inversion of three-beam intensities for scalar scattering by a general centrosymmetric crystal. Acta Cryst*. A**36**, 737–738.

International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited E. Prince, 3rd ed. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (2005). Vol. A. *Space-Group Symmetry*, edited by Th. Hahn, 5th ed. Heidelberg: Springer.

International Tables for X-ray Crystallography (1952). Vol. I. *Symmetry Groups*. Birmingham: Kynoch Press.

Ishizuka, K. (1982). *Translation symmetries in convergent-beam electron-diffraction*. *Ultramicroscopy*, **9**, 255–258.

Ishizuka, K., Miyazaki, M. & Uyeda, N. (1982). *Improvement of electron microscope images by the direct phasing method. Acta Cryst*. A**38**, 408–413.

Jackson, J. I., Meyer, C. H., Nishimura, D. G. & Macovski, A. (1991). *Selection of a convolution function for Fourier inversion using gridding*. *IEEE Trans. Med. Imaging*, **10**, 473–478.

Janner, A. & Janssen, T. (1980*a*). *Symmetry of incommensurate crystal phases. I. Commensurate basic structures*. *Acta Cryst*. A**36**, 399–408.

Janner, A. & Janssen, T. (1980*b*). *Symmetry of incommensurate crystal phases. II. Incommensurate basic structure*. *Acta Cryst*. A**36**, 408–415.

Jap, B. K. & Glaeser, R. M. (1980). *The scattering of high-energy electrons. II. Quantitative validity domains of the single-scattering approximations for organic crystals. Acta Cryst*. A**36**, 57–67.

Jap, B. K., Walian, P. J. & Gehring, K. (1991). *Structural architecture of an outer membrane channel as determined by electron crystallography. Nature (London)*, **350**, 167–170.

Jarić, M. Y. (1988). Editor. *Introduction to Quasicrystals*, Vol. 1. New York: Academic Press.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Improved methods for building protein models in electron density maps and the location of errors in these models*. *Acta Cryst.* A**47**, 110–119.

Joyeux, L. & Penczek, P. A. (2002). *Efficiency of 2D alignment methods*. *Ultramicroscopy*, **92**, 33–46.

Kaczmarz, S. (1993). *Approximate solutions of systems of linear equations (Reprint of Kaczmarz, S., Angenäherte Auflösung von Systemen linearer Gleichunger, Bulletin International de l'Academie Polonaise des Sciences. Lett A, 355–357, 1937)*. *Int. J. Control*, **57**, 1269–1271.

Kambe, K. (1982). *Visualization of Bloch waves of high energy electrons in high resolution electron microscopy. Ultramicroscopy*, **10**, 223–228.

Karle, J. & Hauptman, H. (1956). *A theory of phase determination for the four types of non-centrosymmetric space groups* 1*P*222, 2*P*22, $3P_12$, $3P_22$. *Acta Cryst*. **9**, 635–651.

Kessel, M., Radermacher, M. & Frank, J. (1985). *The structure of the stalk surface layer of a brine pond microorganism: correlation averaging applied to a double layered lattice structure*. *J. Microsc.* **139**, 63–74.

Kirkland, E. J., Siegel, B. M., Uyeda, N. & Fujiyoshi, Y. (1980). *Digital reconstruction of bright field phase contrast images from high resolution electron micrographs. Ultramicroscopy*, **5**, 479–503.

Kiselev, N. A., Lerner, F. Ya. & Livanova, N. B. (1971). *Electron microscopy of muscle phosphorylase B. J. Mol. Biol*. **62**, 537–549.

Klug, A. & Berger, J. E. (1964). *An optical method for the analysis of periodicities in electron micrographs and some observations on the mechanism of negative staining. J. Mol. Biol*. **10**, 565–569.

Klug, A. & DeRosier, D. J. (1966). *Optical filtering of electron micrographs: reconstruction of one-sided images. Nature (London)*, **212**, 29–32.

Kossel, W. & Möllenstedt, G. (1938). *Electron interference in a convergent beam. Nature (London)*, **26**, 660.

Kosykh, V. P., Pustovskikh, A. I., Kirichuk, V. S., Kühne, T., Orlova, E. V., Tsuprun, V. L. & Kiselev, N. A. (1983). *Use of digital storage methods*

to recover images of monocrystalline layers of virus particles. Sov. Phys. Crystallogr. **28**, 637–643.
Krämer, S., Mayer, J., Witt, C., Weikenmeier, A. & Rühle, M. (2000). *Ultramicroscopy*, **81**, 245–262.
Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. (1994). *Atomic model of plant light-harvesting complex by electron crystallography. Nature (London)*, **367**, 614–621.
Kutateladze, T. G., Ogburn, K. D., Watson, W. T., de Beer, T., Emr, S. D., Burd, C. G. & Overduin, M. (1999). *Phosphatidylinositol 3-phosphate recognition by the FYVE domain. Mol. Cell*, **3**, 805–811.
Kuwabara, S. (1978). *Nearly aberration-free crystal images in high voltage electron microscopy. J. Electron Microsc.* **27**, 161–169.
Langer, R., Frank, J., Feltynowski, A. & Hoppe, W. (1970). *Anwendung des Bilddifferenzverfahrens auf die Untersuchung von Strukturänderungen dünner Kohlefolien bei Elektronenbestrahlung. Ber. Bunsenges Phys. Chem.* **74**, 1120–1126.
Langs, D. A. & DeTitta, G. T. (1975). *A flexible and rapid phase determination and refinement procedure. Acta Cryst.* A**31**, S16.
Lanzavecchia, S. & Bellon, P. L. (1994). *A moving window Shannon reconstruction for image interpolation. J. Visual Comm. Image Repres.* **3**, 255–264.
Lanzavecchia, S. & Bellon, P. L. (1998). *Fast computation of 3D Radon transform via a direct Fourier method. Bioinformatics*, **14**, 212–216.
Lanzavecchia, S., Bellon, P. L. & Scatturin, V. (1993). *SPARK, a kernel of software programs for spatial reconstruction in electron microscopy. J. Microsc.* **171**, 255–266.
Lanzavecchia, S., Tosoni, L. & Bellon, P. L. (1996). *Fast sinogram computation and the sinogram-based alignment of images. Comp. Appl. Bio. Sci.* **12**, 531–537.
Lata, K. R., Penczek, P. & Frank, J. (1995). *Automatic particle picking from electron micrographs. Ultramicroscopy*, **58**, 381–391.
Laurette, I., Zeng, G. L., Welch, A., Christian, P. E. & Gullberg, G. T. (2000). *A three-dimensional ray-driven attenuation, scatter and geometric response correction technique for SPECT in inhomogeneous media. Phys. Med. Biol.* **45**, 3459–3480.
Lawrence, M. C. (1992). *Least-squares method of alignment using markers*. In *Electron Tomography*, edited by J. Frank, pp. 197–204. New York: Plenum Press.
Lawton, J. A. & Prasad, B. V. V. (1996). *Automated software package for icosahedral virus reconstruction. J. Struct. Biol.* **116**, 209–215.
Levine, D. & Steinhardt, P. J. (1984). *Quasicrystals – a new class of ordered structures. Phys. Rev. Lett.* **53**, 2477–2480.
Lewitt, R. M. (1990). *Multidimensional digital image representations using generalized Kaiser–Bessel window functions. J. Opt. Soc. Am. A*, **7**, 1834–1846.
Lewitt, R. M. (1992). *Alternatives to voxels for image representation in iterative reconstruction algorithms. Phys. Med. Biol.* **37**, 705–716.
Li, D. X. & Hovmöller, S. (1988). *The crystal structure of $Na_3Nb_{12}O_{31}F$ determined by HREM and image processing. J. Solid State Chem.* **73**, 5–10.
Li, F. H. (1991). *Crystal structures from high-resolution electron microscopy*. In *Electron Crystallography of Organic Molecules*, edited by J. R. Fryer & D. L. Dorset, pp. 153–167. Dordrecht: Kluwer Academic Publishers.
Liang, Y. Y., Ke, E. Y. & Zhou, Z. H. (2002). *IMIRS: a high-resolution 3D reconstruction package integrated with a relational image database. J. Struct. Biol.* **137**, 292–304.
Liebman, G. (1955). *A unified representation of magnetic electron lens properties. Proc. Phys. Soc. London Sect. B*, **68**, 737–745.
Liu, Y.-W., Fan, H.-F. & Zheng, C.-D. (1988). *Image processing in high-resolution electron microscopy using the direct method. III. Structure-factor extrapolation. Acta Cryst.* A**44**, 61–63.
Lobachev, A. N. & Vainshtein, B. K. (1961). *An electron diffraction study of urea. Sov. Phys. Crystallogr.* **6**, 313–317.
Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). *EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol.* **128**, 82–97.
Lynch, D. F. & Moodie, A. F. (1972). *Numerical evaluation of low energy electron diffraction intensity. I. The perfect crystal with no upper layer lines and no absorption. Surf. Sci.* **32**, 422–438.
Lynch, D. F., Moodie, A. F. & O'Keefe, M. A. (1975). *n-Beam lattice images. V. The use of the charge-density approximation in the interpretation of lattice images. Acta Cryst.* A**31**, 300–307.
Mackay, A. L. (1982). *Crystallography and the Penrose pattern. Physica (Utrecht)*, **114A**, 609–613.
McLachlan, D. (1958). *Crystal structure and information theory. Proc. Natl Acad. Sci. USA*, **44**, 948–956.
Mallick, S. P., Carragher, B., Potter, C. S. & Kriegman, D. J. (2005). *ACE: automated CTF estimation. Ultramicroscopy*, **104**, 8–29.
Mao, Y., Nickitenko, A., Duan, X., Lloyd, T. E., Wu, M. N., Bellen, H. & Quiocho, F. A. (2000). *Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. Cell*, **100**, 447–456.
Marabini, R. & Carazo, J. M. (1996). *On a new computationally fast image invariant based on bispectral projections. Pattern Recognit. Lett.* **17**, 959–967.
Marabini, R., Herman, G. T. & Carazo, J. M. (1998). *3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). Ultramicroscopy*, **72**, 53–65.
Markham, R., Frey, S. & Hills, G. J. (1963). *Methods for the enhancement of image detail and accentuation of structure in electron microscopy. Virology*, **20**, 88–102.
Matej, S. & Lewitt, R. M. (1996). *Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. IEEE Trans. Med. Imaging*, **15**, 68–78.
Matsuda, T., Tonomura, A. & Komada, T. (1978). *Observation of lattice images with a field emission electron microscope. Jpn. J. Appl. Phys.* **17**, 2073–2074.
Midgley, P. A., Sleight, M. E. & Vincent, R. (1996). *The structure of a metastable Au–Sn phase determined by convergent beam electron diffraction. J. Solid State Chem.* **124**, 132–142.
Mindell, J. A. & Grigorieff, N. (2003). *Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol.* **142**, 334–347.
Ming, D. M., Kong, Y. F., Lambert, M. A., Huang, Z. & Ma, J. P. (2002). *How to describe the movement of protein without amino acids sequence and coordinates. Proc. Natl Acad. Sci. USA*, **13**, 8620–8625.
Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C. L. III, Ban, N. & Frank, J. (2005). *Structure of the E. coli protein-conducting channel bound to a translating ribosome. Nature (London)*, **438**, 318–324.
Miyake, S., Takagi, S. & Fujimoto, F. (1960). *The extinction rule for reflexions in symmetrical spot patterns of electron diffraction by crystals. Acta Cryst.* **13**, 360–361.
Mo, Y. D., Cheng, T. Z., Fan, H. F., Li, J. Q., Sha, B. D., Zheng, C. D., Li, F. H. & Zhao, Z. X. (1992). *Structural features of the incommensurate modulation in the Pb-doped Bi-2223 high-T_c phase by defect method electron diffraction analysis. Supercond. Sci. Technol.* **5**, 69–72.
Moodie, A. F. (1965). *Some structural implications of n-beam interactions*. International Conference on Electron Diffraction and Crystal Defects, Melbourne, Australia, paper ID-1.
Morniroli, J. P. (2002). *Large-Angle Convergent Beam Electron Diffraction*. Paris: French Society of Microscopy.
Moss, B. & Dorset, D. L. (1982). *Effect of crystal bending on direct phasing of electron diffraction data from cytosine. Acta Cryst.* A**38**, 207–211.
Mouche, F., Boisset, N. & Penczek, P. A. (2001). *Lumbricus terrestris hemoglobin – the architecture of linker chains and structural variation of the central toroid. J. Struct. Biol.* **133**, 176–192.
Nagasawa, T. (1983). Master of Physics Thesis, Tohoku University, Japan.
Natterer, F. & Wübbeling, F. (2001). *Mathematical Methods in Image Reconstruction*. Philadelphia: SIAM.
Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. (2000). *The structural basis of ribosome activity in peptide bond synthesis. Science*, **289**, 920–930.
Nowacki, W. (1967). *Crystal Data*. ACA Monograph No. 6. Washington: American Crystallographic Association.
Ogawa, T. (1985). *On the structure of a quasicrystal 3-dimensional Penrose transformation. J. Phys. Soc. Jpn*, **54**, 3205–3208.
Ogawa, T., Moriguchi, S., Isoda, S. & Kobayashi, T. (1994). *Application of an imaging plate to electron crystallography at atomic resolution. Polymer*, **35**, 1132–1136.
Ogata, Y., Tsuda, K., Akishige, Y. & Tanaka, M. (2004). *Refinement of the crystal structural parameters of the intermediate phase of h-$BaTiO_3$ using convergent-beam electron diffraction. Acta Cryst.* A**60**, 525–531.
Okabe, A., Boots, B., Sugihara, K. & Chiu, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York: John Wiley & Sons.

Orlov, S. S. (1976). *Theory of three-dimensional reconstruction 1. Conditions for a complete set of projections*. *Sov. Phys. Crystallogr.* **20**, 312–314.

O'Sullivan, J. D. (1985). *A fast sinc function gridding algorithm for Fourier inversion in computer tomography*. *IEEE Trans. Med. Imaging*, **4**, 200–207.

Ottensmeyer, F. P., Andrews, J. W., Basett-Jones, D. P., Chan, A. S. & Hewitt, J. (1977). *Signal to noise enhancement in dark field electron micrographs of vasopressin: filtering of arrays of images in reciprocal space. J. Microsc*. **109**, 256–268.

Pan, M. & Crozier, P. A. (1993). *Quantitative imaging and diffraction of zeolites using a slow-scan CCD camera. Ultramicroscopy*, **52**, 487–498.

Parlett, B. N. (1980). *A new look at the Lanczos-algorithm for solving symmetric-systems of linear-equations*. *Linear Algebr. Its Appl.* **29**, 323–346.

Paul, D., Patwardhan, A., Squire, J. M. & Morris, E. P. (2004) *Single particle analysis of filamentous and highly elongated macromolecular assemblies*. *J. Struct. Biol.* **148**, 236–250.

Penczek, P. (1998). *Measures of resolution using Fourier shell correlation*. *J. Mol. Biol.* **280**, 115–116.

Penczek, P., Ban, N., Grassucci, R. A., Agrawal, R. K. & Frank, J. (1999). *Haloarcula marismortui 50S subunit – complementarity of electron microscopy and X-ray crystallographic information*. *J. Struct. Biol.* **128**, 44–50.

Penczek, P., Marko, M., Buttle, K. & Frank, J. (1995). *Double-tilt electron tomography*. *Ultramicroscopy*, **60**, 393–410.

Penczek, P., Radermacher, M. & Frank, J. (1992). *Three-dimensional reconstruction of single particles embedded in ice*. *Ultramicroscopy*, **40**, 33–53.

Penczek, P. A. (2002). *Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms*. *J. Struct. Biol.* **138**, 34–46.

Penczek, P. A., Chao, Y., Frank, J. & Spahn, C. M. T. (2006). *Estimation of variance in single particle reconstruction using the bootstrap technique*. *J. Struct. Biol.* **154**, 168–183.

Penczek, P. A. & Frank, J. (2006). *Resolution in electron tomography*. In *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*, 2nd ed., edited by J. Frank, pp. 307–330. Berlin: Springer.

Penczek, P. A., Frank, J. & Spahn, C. M. T. (2006*a*). *A method of focused classification, based on the bootstrap 3-D variance analysis, and its application to EF-G-dependent translocation*. *J. Struct. Biol.* **154**, 184–194.

Penczek, P. A., Frank, J. & Spahn, C. M. T. (2006*b*). *Conformational analysis of macromolecules analyzed by cryo-electron microscopy*. In *Microscopy and Microanalysis*, edited by P. Kotula, M. Marko, J.-H. Scott *et al.*, p. CD386. Chicago: Cambridge University Press.

Penczek, P. A., Grassucci, R. A. & Frank, J. (1994). *The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles*. *Ultramicroscopy*, **53**, 251–270.

Penczek, P. A., Renka, R. & Schomberg, H. (2004). *Gridding-based direct Fourier inversion of the three-dimensional ray transform*. *J. Opt. Soc. Am. A*, **21**, 499–509.

Penczek, P. A., Zhu, J. & Frank, J. (1996). *A common-lines based method for determining orientations for N > 3 particle projections simultaneously*. *Ultramicroscopy*, **63**, 205–218.

Penczek, P. A., Zhu, J., Schröder, R. & Frank, J. (1997). *Three-dimensional reconstruction with contrast transfer function compensation from defocus series*. *Scan. Microsc. Suppl.* **11**, 1–10.

Penrose, R. (1974). *The role of aesthetics in pure and applied mathematical research*. *Bull. Inst. Math. Appl.* **10**, 266–271.

Pérez, S. & Chanzy, H. (1989). *Electron crystallography of linear polysaccharides. J. Electron Microsc. Tech*. **11**, 280–285.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *UCSF Chimera – a visualization system for exploratory research and analysis*. *J. Comput. Chem.* **25**, 1605–1612.

Picture Processing and Digital Filtering (1975). Edited by T. S. Huang. Berlin: Springer-Verlag.

Pinsker, Z. G. (1953). *Electron Diffraction*. London: Butterworth.

Pinsker, Z. G., Zvyagin, B. B. & Imamov, R. M. (1981). *Principal results of electron-diffraction structural investigations. Sov. Phys. Crystallogr*. **26**, 669–674.

Pogany, A. P. & Turner, P. S. (1968). *Reciprocity in electron diffraction and microscopy. Acta Cryst*. A**24**, 103–109.

Portier, R. & Gratias, D. (1981). *Diffraction symmetries for elastic scattering*. In *Electron Microscopy and Analysis. Inst. Phys. Conf. Ser*. No. 61, pp. 275–278. Bristol, London: Institute of Physics.

Powell, M. J. D. (1973). *On search directions for minimization algorithm*. *Math. Program.* **4**, 193–201.

Provencher, S. W. & Vogel, R. H. (1988). *Three-dimensional reconstruction from electron micrographs of disordered specimens. I. Method*. *Ultramicroscopy*, **25**, 209–221.

Pullan, L., Mullapudi, S., Huang, Z., Baldwin, P. R., Chin, C., Sun, W., Tsujimoto, S., Kolodziej, S., Stoops, J. K., Lee, J. C., Waxham, M. N., Bean, A. J. & Penczek, P. A. (2006). *The endosome-associated protein Hrs is hexameric and controls cargo sorting as a 'master molecule'*. *Structure*, **14**, 661–671.

Radermacher, M. (1992). *Weighted back-projection methods*. In *Electron Tomography*, edited by J. Frank, pp. 91–115. New York: Plenum.

Radermacher, M. (1994). *Three-dimensional reconstruction from random projections: orientational alignment via Radon transforms*. *Ultramicroscopy*, **53**, 121–136.

Radermacher, M. (2000). *Three-dimensional reconstruction of single particles in electron microscopy*. In *Image Analysis: Methods and Applications*, edited by D.-P. Häder, pp. 295–328. Boca Raton: CRC Press.

Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. (1986). *A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of E. coli*. *J. Microsc.* **141**, RP1–RP2.

Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. (1987). *Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli*. *J. Microsc.* **146**, 113–136.

Radi, G. (1970). *Complex lattice potentials in electron diffraction calculated for a number of crystals. Acta Cryst*. A**26**, 41–56.

Renka, R. J. (1997). *Algorithm 772. STRIPACK: Delaunay triangulation and Voronoi diagram on the surface of a sphere*. *ACM Trans. Math. Software*, **23**, 416–434.

Revol, J. F. (1991). *Electron crystallography of radiation-sensitive polymer crystals*. In *Electron Crystallography of Organic Molecules*, edited by J. R. Fryer & D. L. Dorset, pp. 169–187. Dordrecht: Kluwer Academic Publishers.

Revol, J. F. & Manley, R. St. J. (1986). *Lattice imaging in polyethylene single crystals*. *J. Mater. Sci. Lett*. **5**, 249–251.

Rez, P. (1978). In *Electron Diffraction 1927–1977*, edited by P. J. Dobson, J. B. Pendry & C. J. Humphreys, pp. 61–67. *Inst. Phys. Conf. Ser*. No. 41. Bristol, London: Institute of Physics.

Roseman, A. M. (2003). *Particle finding in electron micrographs using a fast local correlation algorithm*. *Ultramicroscopy*, **94**, 225–236.

Roseman, A. M., Chen, S., White, H., Braig, K. & Saibil, H. R. (1996). *The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL*. *Cell*, **87**, 241–251.

Rossouw, C. J., Gibson, M. A. & Forwood, C. T. (1996). *Dynamical electron diffraction analysis of lattice parameters, Debye–Waller factors and order in Ti–Al and Ti–Ga alloys*. *Ultramicroscopy,* **66**, 193–209.

Rozenfeld, A. (1969). *Picture Processing by Computer*. New York: Academic Press.

Ruiz, T., Mechin, I., Bar, J., Rypniewski, W., Kopperschlager, G. & Radermacher, M. (2003). *The 10.8-A structure of Saccharomyces cerevisiae phosphofructokinase determined by cryoelectron microscopy: localization of the putative fructose 6-phosphate binding sites*. *J. Struct. Biol.* **143**, 124–134.

Saad, A., Ludtke, S. J., Jakana, J., Rixon, F. J., Tsuruta, H. & Chiu, W. (2001). *Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination*. *J. Struct. Biol.* **133**, 32–42.

Saito, M., Tanaka, M., Tsai, A. P., Inoue, A. & Masumoto, T. (1992). *Space group determination of decagonal quasi-crystals of an $Al_{70}Ni_{15}Fe_{15}$ alloy using convergent-beam electron-diffraction*. *Jpn. J. Appl. Phys*. **31**, L109–L112.

Saitoh, K., Tanaka, M. & Tsai, A. P. (2001). *Structural study of an $Al_{73}Ni_{22}Fe_5$ decagonal quasicrystal by high-angle annular dark-field scanning transmission electron microscopy*. *J. Electron Microsc.* **50**, 197–203.

Saitoh, K., Tsuda, K., Tanaka, M., Kaneko, K. & Tsai, A. P. (1997). *Structural study of an $Al_{72}Ni_{20}Co_8$ decagonal quasicrystal using the high-angle annular dark-field method*. *Jpn. J. Appl. Phys.* **36**, L1400–L1402.

Saitoh, K., Tsuda, K., Tanaka, M. & Tsai, A. P. (1999). *Structural study of an $Al_{70}Ni_{15}Fe_{15}$ decagonal quasicrystal using high-angle annular dark-field scanning transmission electron microscopy*. *Jpn. J. Appl. Phys.* **38**, L671–L674.

Saitoh, K., Tsuda, K., Terauchi, M. & Tanaka, M. (2001). *Distinction between space groups having principal rotation and screw axes, which are combined with twofold rotation axes, using the coherent convergent-beam electron diffraction method*. *Acta Cryst.* A**57**, 219–230.

Sander, B., Golas, M. M. & Stark, H. (2003). *Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra*. *J. Struct. Biol.* **142**, 392–401.

Saunders, M., Bird, D. M., Zaluzee, N. J., Burgess, W. G., Preston, A. R. & Humphreys, C. J. (1995). *Measurement of the low-order structure factors for silicon from zone-axis CBED pattern*. *Ultramicroscopy*, **60**, 311–323.

Saxton, W. O. (1978). *Computer Techniques for Image Processing of Electron Microscopy*. New York: Academic Press.

Saxton, W. O. & Baumeister, W. (1982). *The correlation averaging of a regularly arranged bacterial envelope protein*. *J. Microsc.* **127**, 127–138.

Sayre, D. (1952). *The squaring method: a new method for phase determination. Acta Cryst*. **5**, 60–65.

Sayre, D. (1980). *Phase extension and refinement using convolutional and related equation systems*. In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 271–286. New York: Plenum Press.

Scaringe, R. P. (1992). *Crystallography in two dimensions: comparison of theory and experiment for molecular layers. Trans. Am. Crystallogr. Assoc*. **28**, 11–23.

Schatz, M. & van Heel, M. (1990). *Invariant classification of molecular views in electron micrographs*. *Ultramicroscopy*, **32**, 255–264.

Schatz, M. & van Heel, M. (1992). *Invariant recognition of molecular projections in vitreous ice preparations*. *Ultramicroscopy*, **45**, 15–22.

Scherzer, O. (1949). *The theoretical resolution limit of the electron microscope. J. Appl. Phys*. **20**, 20–29.

Schiske, P. (1968). *Zur Frage der Bildrekonstruktion durch Fokusreihen*. 1 *Y Eur. Reg. Conf. Electron Microsc. Rome*, **1**, 145–146.

Schomberg, H. (2002). *Notes on direct and gridding-based Fourier inversion methods*. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Washington, DC, edited by M. Unser & Z. P. Liang, pp. 645–648.

Schomberg, H. & Timmer, J. (1995). *The gridding method for image reconstruction by Fourier transformation*. *IEEE Trans. Med. Imaging*, **14**, 596–607.

Schüler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P. A., Westhof, E. & Spahn, C. M. T. (2006). *Structure of the ribosome-bound cricket paralysis virus IRES RNA*. *Nature Struct. Mol. Biol.* **13**, 1092–1096.

Sezan, M. I. (1992). *An overview of convex projections theory and its application to image recovery problems*. *Ultramicroscopy*, **40**, 55–67.

Sezan, M. I. & Stark, H. (1982). *Image restoration by the method of convex projections. II. Applications and numerical results*. *IEEE Trans. Med. Imaging*, **1**, 95–101.

Sha, B.-D., Fan, H.-F. & Li, F.-H. (1993). *Correction for the dynamical electron diffraction effect in crystal structure analysis. Acta Cryst*. A**49**, 877–880.

Shannon, C. E. (1949). *Communication in the presence of noise*. *Proc. Inst. Radio Eng.* **37**, 10–21.

Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. (1984). *Metallic phase with long-range orientational order and no translational symmetry. Phys. Rev. Lett*. **53**, 1951–1953.

Shoemaker, V. & Glauber, R. (1952). *The Born approximation in electron diffraction. Nature (London)*, **170**, 290–291.

Sigworth, F. J. (2004). *Classical detection theory and the cryo-EM particle selection problem*. *J. Struct. Biol.* **145**, 111–122.

Sorzano, C. O. S., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H. W., Carazo, J. M. & Pascual-Montano, A. (2004). *XMIPP: a new generation of an open-source image processing package for electron microscopy*. *J. Struct. Biol.* **148**, 194–204.

Spahn, C. M. T., Penczek, P., Leith, A. & Frank, J. (2000). *A method for differentiating proteins from nucleic acids in intermediate-resolution density maps: cryo-electron microscopy defines the quaternary structure of the Escherichia coli 70S ribosome*. *Struct. Fold. Des.* **8**, 937–948.

Spence, J. C. H. (1993). *On the accurate measurement of structure-factor amplitudes and phases by electron diffraction*. *Acta Cryst*. A**49**, 231–260.

Spence, J. C. H. (2003). *High-Resolution Electron Microscopy*. New York: Oxford University Press.

Spence, J. C. H. & Cowley, J. M. (1978). *Lattice imaging in STEM*. *Optik*, **50**, 129–142.

Spence, J. C. H., O'Keefe, M. A. & Kolar, H. (1977). *Image interpretation in crystalline germanium. Optik (Stuttgart)*, **49**, 307–323.

Spence, J. C. H. & Zuo, J. M. (1992). *Electron Microdiffraction*. New York: Plenum Press.

Steeds, J. W., Bird, D. M., Eaglesham, D. J., Mckernen, S., Vincent, R. & Withers, R. L. (1985). *Study of modulated structures by transmission electron microscopy*. *Ultramicroscopy*, **18**, 97–110.

Steeds, J. W. & Evans, N. S. (1980). *Practical examples of point and space group determination in convergent beam diffraction. Proc. Electron Microsc. Soc. Am*. pp. 188–191.

Steeds, J. W., Rackham, G. M. & Shannon, M. D. (1978). *On the observation of dynamically forbidden lines in two and three dimensional electron diffraction*. In *Electron Diffraction 1927–1977. Inst. Phys. Conf. Ser*. No. 41, pp. 135–139.

Steinkilberg, M. & Schramm, H. J. (1980). *Eine verbesserte Drehkorrelations Methode für die Strukturbestimmung biologischer Macromoleküle durch Mittelung elektronenmikroskopischer Bilder. Hoppe–Seyler's Z. Physiol. Chem*. **361**, 1363–1369.

Stereochemical Applications of Gas-Phase Electron Diffraction (1988). Part A, edited by I. Hargittai & M. Hargittai. New York: VCH.

Steurer, W. (2004). *Twenty years of structure research on quasicrystals. Part 1. Pentagonal, octagonal, decagonal and dodecagonal quasicrystals. Z. Kristallogr.* **219**, 391–446.

Taftø, T. & Spence, J. C. H. (1982). *A simple method for the determination of structure-factor phase relationships and crystal polarity using electron diffraction*. *J. Appl. Cryst.* **15**, 60–64.

Tanaka, M. (1986). *Conventional transmission-electron-microscopy techniques in convergent-beam electron diffraction*. *J. Electron Microsc.* **35**, 314–323.

Tanaka, M. (1989). *Symmetry analysis*. *J. Electron Microsc. Tech.* **13**, 27–39.

Tanaka, M., Saito, R. & Sekii, H. (1983). *Point-group determination by convergent-beam electron diffraction*. *Acta Cryst.* A**39**, 357–368.

Tanaka, M., Saito, P., Ueno, K. & Harada, Y. (1980). *Large angle convergent-beam electron diffraction. J. Electron. Microsc*. **29**, 408–412.

Tanaka, M., Sekii, H. & Nagasawa, T. (1983). *Space group determination by dynamic extinction in convergent beam electron diffraction. Acta Cryst*. A**39**, 825–837.

Tanaka, M., Takayoshi, H., Ishida, M. & Endoh, Y. (1985). *Crystal chirality and helicity of the helical spin-density wave in MnSi. 1. Convergent-beam electron diffraction*. *J. Phys. Soc. Jpn*, **54**, 2970–2974.

Tanaka, M. & Terauchi, M. (1985). *Convergent-Beam Electron Diffraction*. Tokyo: JEOL Ltd.

Tanaka, M., Terauchi, M. & Kaneyama, T. (1988). *Convergent-Beam Electron Diffraction II*. Tokyo: JEOL Ltd.

Tanaka, M., Terauchi, M. & Sekii, H. (1987). *Observation of dynamic extinction due to a glide plane perpendicular to an incident beam by convergent-beam electron diffraction*. *Ultramicroscopy*, **21**, 245–250.

Tanaka, M., Terauchi, M., Suzuki, S., Hiraga, K. & Hirabayashi, M. (1987). *Convergent-beam and small-area-parallel-beam electron diffraction patterns of icosahedral quasicrystals of an $Al_{74}Mn_{20}Si_6$ alloy*. *Acta Cryst.* B**43**, 494–501.

Tanaka, M., Terauchi, M. & Tsuda, K. (1994). *Convergent-Beam Electron Diffraction III*. Tokyo: JEOL–Maruzen.

Tanaka, M., Terauchi, M., Tsuda, K. & Saitoh, K. (2002). *Convergent-Beam Electron Diffraction IV*, p. 13. Tokyo: JEOL.

Tanaka, M. & Tsuda, K. (1990). *Determination of positional parameters by convergent-beam electron diffraction*. *Proc. XIIth Int. Congr. Electron Microsc.*, Seattle, edited by L. D. Peachy & D. B. Williams, Vol. 2, pp. 518–519. San Francisco: San Francisco Press.

Tanaka, M. & Tsuda, K. (1991). *Microbeam Analysis*, edited by D. G. Howitt, pp. 145–146. San Francisco: San Francisco Press.

Tanaka, M., Tsuda, K. & Saitoh, K. (1996). *Convergent-beam electron diffraction and electron microscope studies of decagonal quasicrystals*. *Sci. Rep. RITU*, **A42**, 199–205.

Tanaka, M., Tsuda, K., Terauchi, M., Fujiwara, A., Tsai, A. P., Inoue, A. & Masumoto, T. (1993). *Electron diffraction and electron microscope study on decagonal quasicrytals of Al–Ni–Fe alloys*. *J. Non-Cryst. Solids*, **153&154**, 98–102.

Terauchi, M., Takahashi, M. & Tanaka, M. (1994). *Convergent-beam electron diffraction study of incommensurately modulated crystals. II. (3+1)-dimensional space groups*. *Acta Cryst*. A**50**, 566–574.

Terauchi, M. & Tanaka, M. (1993). *Convergent-beam electron diffraction study of incommensurately modulated crystals. I. (3+1)-dimensional point groups*. *Acta Cryst*. A**49**, 722–729.

Thon, F. (1966). *On the defocusing dependence of phase contrast in electron microscopical images. Z. Naturforsch. Teil A*, **21**, 476–478.

Tinnappel, A. (1975). PhD Thesis, Technische Universität Berlin, Germany.

Tivol, W. F., Dorset, D. L., McCourt, M. P. & Turner, J. N. (1993). *Voltage-dependent effect on dynamical scattering and the electron diffraction structure analysis of organic crystals: copper perchlorophthalocyanine. Microsc. Soc. Am. Bull*. **23**, 91–98.

Tournaire, M. (1962). *Recent developments of the matrical and semi-reciprocal formulation in the field of dynamical theory. J. Phys. Soc. Jpn*, **17**, Suppl. B11, 98–100.

Tsai, A. P. (2003). *'Back to the future' – An account of the discovery stable quasicrystals. Acc. Chem. Res.* **36**, 31–38.

Tsai, A. P., Inoue, A. & Masumoto, T. (1987). *A stable quasi-crystal in Al–Cu–Fe system*. *Jpn. J. Appl. Phys. Lett.* **26**, L1505–L1507.

Tsai, A. P., Inoue, A. & Masumoto, T. (1989*a*). *A stable decagonal quasicrystal in the Al–Cu–Co system*. *Mater. Trans. Jpn. Inst. Met.* **30**, 300–304.

Tsai, A. P., Inoue, A. & Masumoto, T. (1989*b*). *Stable decagonal Al–Co–Ni and Al–Co–Cu quasicrystals*. *Mater. Trans. Jpn. Inst. Met.* **30**, 463–473.

Tsai, A. P., Inoue, A. & Masumoto, T. (1989*c*). *New decagonal Al–Ni–Fe and Al–Ni–Co alloys prepared by liquid quenching*. *Mater. Trans. Jpn. Inst. Met.* **30**, 150–154.

Tsipursky, S. I. & Drits, V. A. (1977). *Efficiency of electronometric intensity registration at electron diffraction structural studies. Izv. Akad. Nauk SSSR Ser. Fiz.* **41**, 2263–2271.

Tsuda, K., Ogata, Y., Takagi, K., Hashimoto, T. & Tanaka, M. (2002). *Refinement of crystal structural parameters and charge density using convergent-beam electron diffraction – the rhombohedral phase of $LaCrO_3$*. *Acta Cryst*. A**58**, 514–525.

Tsuda, K., Saito, M., Terauchi, M., Tanaka, M., Tsai, A. P., Inoue, A. & Masumoto, K. (1993). *Electron microscope study of decagonal quasicrystals of $Al_{70}Ni_{15}Fe_{15}$*. *Jpn. J. Appl. Phys.* **32**, 129–134.

Tsuda, K., Saitoh, K., Terauchi, M., Tanaka, M. & Goodman, P. (2000). *Distinction of space groups (I23 and $I2_13$) and (I222 and $I2_12_12_1$) using coherent convergent-beam electron diffraction*. *Acta Cryst.* A**56**, 359–369.

Tsuda, K. & Tanaka, M. (1995). *Refinement of crystal structure parameters using convergent-beam electron diffraction: the low-temperature phase of $SrTiO_3$*. *Acta Cryst*. A**51**, 7–19.

Tsuda, K. & Tanaka, M. (1999). *Refinement of crystal structural parameters using two-dimensional energy-filtered CBED patterns*. *Acta Cryst*. A**55**, 939–954.

Tsuji, M. (1989). *Electron microscopy*. In *Comprehensive Polymer Science*, Vol. 1. *Polymer Characterization*, edited by C. Booth & C. Price, pp. 785–840. Oxford: Pergamon Press

Turner, P. S. & Cowley, J. M. (1969). *The effects of n-beam dynamical diffraction on electron diffraction intensities from polycrystalline materials. Acta Cryst*. A**25**, 475–481.

Unser, M., Trus, B. L. & Steven, A. C. (1987). *A new resolution criterion based on spectral signal-to-noise ratios*. *Ultramicroscopy*, **23**, 39–51.

Unwin, P. N. T. & Henderson, R. (1975). *Molecular structure determination by electron microscopy of unstained crystalline specimens. J. Mol. Biol*. **94**, 425–440.

Uyeda, N., Kobayashi, T., Ishizuka, K. & Fujiyoshi, Y. (1978–1979). *High voltage electron microscopy for image discrimination of constituent atoms in crystals and molecules. Chem. Scr*. **14**, 47–61.

Vainshtein, B. K. (1952). *Dependence of electron scattering on the atomic number. Dokl. Akad. Nauk SSSR*, **85**, 1239–1242.

Vainshtein, B. K. (1954). *On the studies of crystal lattice potential by electron diffraction. Tr. Inst. Krist. Akad. Nauk SSSR*, **9**, 259–276.

Vainshtein, B. K. (1955). *Elektronograficheskoe issledovanie diketopiperazina. Zh. Fiz. Khim*. **29**, 327–344.

Vainshtein, B. K. (1956). *Structure Analysis by Electron Diffraction*. Moscow: Akad. Sci. USSR.

Vainshtein, B. K. (1964). *Structure Analysis by Electron Diffraction*. Oxford: Pergamon Press.

Vainshtein, B. K. (1971). *Finding the structure of objects from projections. Sov. Phys. Crystallogr*. **15**, 781–787.

Vainshtein, B. K., D'yakon, I. A. & Ablov, A. V. (1971). *Electron diffraction determination of the structure of copper* DL*-alaninate. Sov. Phys. Dokl*. **15**, 645–647.

Vainshtein, B. K. & Klechkovskaya, V. V. (1993). *Electron diffraction by Langmuir–Blodgett films. Proc. R. Soc. London Ser. A*, **442**, 73–84.

Vilkov, L. V., Mastryukov, V. S. & Sadova, N. I. (1978). *Determination of Geometrical Structure of Free Molecules*. Leningrad: Khimiya.

Vincent, R., Bird, D. M. & Steeds, J. W. (1984*a*). *Structure of AuGeAs determined by convergent-beam electron-diffraction. 1. Derivation of basic structure. Philos. Mag. A*, **50**, 745–763.

Vincent, R., Bird, D. M. & Steeds, J. W. (1984*b*). *Structure of AuGeAs determined by convergent-beam electron-diffraction. 2. Refinement of structural parameters. Philos. Mag. A*, **50**, 765–786.

Vincent, R. & Exelby, D. R. (1991). *Structure of metastable Al–Ge phases determined from HOLZ Patterson transforms. Philos. Mag. Lett*. **63**, 31–38.

Vincent, R. & Exelby, D. R. (1993). *Structure of a metastable Al–Ge phase determined from large angle CBED patterns. Philos. Mag. B*, **68**, 513–528.

Vincent, R. & Midgley, P. A. (1994). *Double conical beam rocking system for measurement of integrated electron diffraction intensities. Ultramicroscopy*, **53**, 271–282.

Vine, W. J., Vincent, R., Spellward, P. & Steeds, J. W. (1992). *Observation of phase contrast in convergent-beam electron diffraction*. *Ultramicroscopy*, **41**, 423–428.

Vogel, R. H. & Provencher, S. W. (1988). *Three-dimensional reconstruction from electron micrographs of disordered specimens. II. Implementation and results*. *Ultramicroscopy*, **25**, 223–239.

Voronova, A. A. & Vainshtein, B. K. (1958). *An electron diffraction study of $CuCl_2{\cdot}3Cu(OH)_2$. Sov. Phys. Crystallogr.* **3**, 445–451.

Walz, J., Typke, D., Nitsch, M., Koster, A. J., Hegerl, R. & Baumeister, W. (1997). *Electron tomography of single ice-embedded macromolecules: three-dimensional alignment and classification*. *J. Struct. Biol.* **120**, 387–395.

Watanabe, D., Uyeda, R. & Kogiso, M. (1968). *An apparent variation of structure factors for electrons with accelerating voltage. An observation through Kikuchi patterns. Acta Cryst*. A**24**, 249–250.

Welch, P. D. (1967). *The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short modified periodograms*. *IEEE Trans. Audio Electroacoust.* **AU-15**, 70–73.

Wenk, H.-R., Downing, K. H., Ho, M.-S. & O'Keefe, M. A. (1992). *3D structure determination from electron-microscope images: electron crystallography of staurolite. Acta Cryst*. A**48**, 700–716.

Wolff, P. M. de (1974). *The pseudo-symmetry of modulated crystal structures*. *Acta Cryst*. A**30**, 777–785.

Wolff, P. M. de (1977). *Symmetry operations for displacively modulated structures*. *Acta Cryst*. A**33**, 493–497.

Wolff, P. M. de, Janssen, T. & Janner, A. (1981). *The superspace groups for incommensurate crystal structures with a one-dimensional modulation. Acta Cryst*. A**37**, 625–636.

Wong, H. C., Chen, J., Mouche, F., Rouiller, I. & Bern, M. (2004). *Model-based particle picking for cryo-electron microscopy*. *J. Struct. Biol.* **145**, 157–167.

Xiang, S.-B., Fan, H.-F., Wu, X.-J., Li, F.-H. & Pan, Q. (1990). *Direct methods in superspace. II. The first application to an unknown incommensurate modulated structure. Acta Cryst*. A**46**, 929–934.

Yamamoto, A. (1982). *Structure factor of modulated crystal structures*. *Acta Cryst.* A**38**, 87–92.

Yamamoto, A. (1996). *Crystallography of quasiperiodic crystals*. *Acta Cryst.* A**52**, 509–560.

Yamamoto, A. & Ishihara, K. N. (1988). *Penrose patterns and related structures. II. Decagonal quasicrystals*. *Acta Cryst.* A**44**, 707–714.

Yamamoto, A., Janssen, T., Janner, A. & de Wolff, P. M. (1985). *A note on the superspace groups for one-dimensionally modulated structures*. *Acta Cryst*. A**41**, 528–530.

Yamamoto, N. (1988). *Electron microscope study of incommensurate phases*. *Kotaibutsui*, **23**, 547.

Yang, C., Ng, E. G. & Penczek, P. A. (2005). *Unified 3-D structure and projection orientation refinement using quasi-Newton algorithm*. *J. Struct. Biol.* **149**, 53–64.

Yao, J.-X. (1981). *On the application of phase relationships to complex structures. XVIII. RANTAN – random MULTAN. Acta Cryst*. A**37**, 642–644.

Yin, Z. H., Zheng, Y. L., Doerschuk, P. C., Natarajan, P. & Johnson, J. E. (2003). *A statistical approach to computer processing of cryo-electron microscope images: virion classification and 3-D reconstruction*. *J. Struct. Biol.* **144**, 24–50.

Yin, Z. Y., Zheng, Y. L. & Doerschuk, P. C. (2001). *An ab initio algorithm for low-resolution 3-D reconstructions from cryoelectron microscopy images*. *J. Struct. Biol.* **133**, 132–142.

Youla, D. C. & Webb, H. (1982). *Image restoration by the method of convex projections. 1. Theory*. *IEEE Trans. Med. Imaging*, **1**, 81–94.

Yuen, C. K. & Fraser, D. (1979). *Digital Spectral Analysis*. Adelaide: CSIRO Pitman.

Zemlin, F., Reuber, E., Beckmann, E., Zeitler, E. & Dorset, D. L. (1985). *Molecular resolution electron micrographs of monolamellar paraffin crystal. Science*, **229**, 461–462.

Zhu, J., Penczek, P. A., Schröder, R. & Frank, J. (1997). *Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome*. *J. Struct. Biol.* **118**, 197–219.

Zhu, Y., Carragher, B., Glaeser, R. M., Fellmann, D., Bajaj, C., Bern, M., Mouche, F., de Haas, F., Hall, R. J., Kriegman, D. J., Ludtke, S. C., Mallick, S. P., Penczek, P. A., Roseman, A. M., Sigworth, F. J., Volkmann, N. & Potter, C. S. (2004). *Automatic particle selection: results of a comparative study*. *J. Struct. Biol.* **145**, 3–14.

Zhukhlistov, A. P., Avilov, A. S., Ferraris, G., Zvyagin, B. B. & Plotnikov, V. P. (1997). *Statistical distribution of hydrogen over three positions in the brucite $Mg(OH)_2$ structure from electron diffractometry data. Crystallogr. Rep.* **42**, 774–777.

Zhukhlistov, A. P. & Zvyagin, B. B. (1998). *Crystal structure of lizardite 1T from electron diffractometry data. Crystallogr. Rep.* **43**, 950–955.

Zuo, J. M., Kim, M., O'Keeffe, M. & Spence, J. C. H. (1999). *Observation of d holes and Cu–Cu bonding in cuprite*. *Nature (London)*, **401**, 49–52.

Zvyagin, B. B. (1967). *Electron-Diffraction Analysis of Clay Mineral Structures*. New York: Plenum.

Zvyagin, B. B., Vrublevskaya, Z. V., Zhukhlistov, A. P., Sidorenko, S. V., Soboleva, A. F. & Fedotov, A. F. (1979). *High-Voltage Electron Diffraction Investigations of Layered Minerals*. Moscow: Nauka.

Zvyagin, B. B., Zhukhlistov, A. P. & Plotnikov, A. P. (1996). *Development of the electron diffractometry of minerals. Structural studies of crystals*. (Coll. Works 75th Anniversary Acad. B. K. Vainshtein.) *Nauka-Physmathlit*, pp. 225–234.

International Tables for Crystallography (2010). Vol. B, Chapter 3.1, pp. 404–409.

3.1. Distances, angles, and their standard uncertainties

By D. E. Sands

3.1.1. Introduction

A crystal structure analysis provides information from which it is possible to compute distances between atoms, angles between interatomic vectors, and the uncertainties in these quantities. In Cartesian coordinate systems, these geometric computations require the Pythagorean theorem and elementary trigonometry. The natural coordinate systems of crystals, though, are determined by symmetry, and only in special cases are the basis vectors (or coordinate axes) of these systems constrained to be of equal lengths or mutually perpendicular.

It is possible, of course, to transform the positional parameters of the atoms to a Cartesian system and perform the subsequent calculations with the transformed coordinates. Along with the coordinates, the transformations must be applied to anisotropic thermal factors, variance–covariance matrices and other important quantities. Moreover, leaving the natural coordinate system of the crystal sacrifices the simplified relationships imposed by translational and point symmetry; for example, if an atom has fractional coordinates x^1, x^2, x^3, an equivalent atom will be at $1 + x^1$, x^2, x^3, *etc.*

Fortunately, formulation of the calculations in generalized rectilinear coordinate systems is straightforward, and readily adapted to computer languages (Section 3.1.12 illustrates the use of Fortran for such calculations). The techniques for these computations are those of tensor analysis, which provides a compact and elegant notation. While an effort will be made to be self-sufficient in this chapter, some proficiency in vector algebra is assumed, and the reader not familiar with the basics of tensor analysis should refer to Chapter 1.1 and Sands (1982*a*).

3.1.2. Scalar product

The scalar product of vectors **u** and **v** is defined as

$$\mathbf{u} \cdot \mathbf{v} = uv \cos\varphi, \tag{3.1.2.1}$$

where u and v are the lengths of the vectors and φ is the angle between them. In terms of components,

$$\mathbf{u} \cdot \mathbf{v} = (u^i \mathbf{a}_i) \cdot (v^j \mathbf{a}_j) \tag{3.1.2.2}$$

$$\mathbf{u} \cdot \mathbf{v} = u^i v^j \mathbf{a}_i \cdot \mathbf{a}_j \tag{3.1.2.3}$$

$$\mathbf{u} \cdot \mathbf{v} = u^i v^j g_{ij}. \tag{3.1.2.4}$$

In all equations in this chapter, the convention is followed that summation is implied over an index that is repeated once as a subscript and once as a superscript in an expression; thus, the right-hand side of (3.1.2.4) implies the sum of nine terms

$$u^1 v^1 g_{11} + u^1 v^2 g_{12} + \ldots + u^3 v^3 g_{33}.$$

The g_{ij} in (3.1.2.4) are the components of the metric tensor [see Chapter 1.1 and Sands (1982*a*)]

$$g_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j. \tag{3.1.2.5}$$

Subscripts are used for quantities that transform the same way as the basis vectors $\mathbf{a}_i$; such quantities are said to transform covariantly. Superscripts denote quantities that transform the same way as coordinates x^i; these quantities are said to transform contravariantly (Sands, 1982*a*).

Equation (3.1.2.4) is in a form convenient for computer evaluation, with indices i and j taking successively all values from 1 to 3. The matrix form of (3.1.2.4) is useful both for symbolic manipulation and for computation,

$$\mathbf{u} \cdot \mathbf{v} = \boldsymbol{u}^T \boldsymbol{g} \boldsymbol{v}, \tag{3.1.2.6}$$

where the superscript italic T following a matrix symbol indicates a transpose. Written out in full, (3.1.2.6) is

$$\mathbf{u} \cdot \mathbf{v} = (u^1 u^2 u^3) \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix}. \tag{3.1.2.7}$$

If $\boldsymbol{u}$ is the column vector with components u^1, u^2, u^3, $\boldsymbol{u}^T$ is the corresponding row vector shown in (3.1.2.7).

3.1.3. Length of a vector

By (3.1.2.1), the scalar product of a vector with itself is

$$\mathbf{v} \cdot \mathbf{v} = (v)^2. \tag{3.1.3.1}$$

The length of **v** is, therefore, given by

$$v = (v^i v^j g_{ij})^{1/2}. \tag{3.1.3.2}$$

Computation of lengths in a generalized rectilinear coordinate system is thus simply a matter of evaluating the double summation $v^i v^j g_{ij}$ and taking the square root.

3.1.4. Angle between two vectors

By (3.1.2.1) and (3.1.2.4), the angle φ between vectors **u** and **v** is given by

$$\varphi = \cos^{-1}[u^i v^j g_{ij}/(uv)]. \tag{3.1.4.1}$$

An even more concise expression of equations such as (3.1.4.1) is possible by making use of the ability of the metric tensor $\boldsymbol{g}$ to convert components from contravariant to covariant (Sands, 1982*a*). Thus,

$$v_i = g_{ij} v^j, \quad u_j = g_{ij} u^i, \tag{3.1.4.2}$$

and (3.1.2.4) may be written succinctly as

$$\mathbf{u} \cdot \mathbf{v} = u^i v_i \tag{3.1.4.3}$$

Copyright © 2010 International Union of Crystallography

or

$$\mathbf{u} \cdot \mathbf{v} = u_i v^i. \tag{3.1.4.4}$$

With this notation, the angle calculation of (3.1.4.1) becomes

$$\varphi = \cos^{-1}[u^i v_i/(uv)] = \cos^{-1}[u_i v^i/(uv)]. \tag{3.1.4.5}$$

The summations in (3.1.4.3), (3.1.4.4) and (3.1.4.5) include only three terms, and are thus equivalent in numerical effort to the computation in a Cartesian system, in which the metric tensor is represented by the unit matrix and there is no numerical distinction between covariant components and contravariant components.

Appreciation of the elegance of tensor formulations may be enhanced by noting that corresponding to the metric tensor $\boldsymbol{g}$ with components g_{ij} there is a contravariant metric tensor $\boldsymbol{g}^*$ with components

$$g^{ij} = \mathbf{a}^i \cdot \mathbf{a}^j. \tag{3.1.4.6}$$

The $\mathbf{a}^i$ are contravariant basis vectors, known to crystallographers as reciprocal axes. Expressions parallel to (3.1.4.2) may be written, in which $\boldsymbol{g}^*$ plays the role of converting covariant components to contravariant components. These tensors thus express mathematically the crystallographic notions of crystal space and reciprocal space [see Chapter 1.1 and Sands (1982*a*)].

3.1.5. Vector product

The scalar product defined in Section 3.1.2 is one multiplicative operation of two vectors that may be defined; another is the vector product, which is denoted as $\mathbf{u} \wedge \mathbf{v}$ (or $\mathbf{u} \times \mathbf{v}$ or $[\mathbf{uv}]$). The vector product of vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as a vector of length $uv \sin\varphi$, where φ is the angle between the vectors, and of direction perpendicular to both $\mathbf{u}$ and $\mathbf{v}$ in the sense that $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} \wedge \mathbf{v}$ form a right-handed system; $\mathbf{u} \wedge \mathbf{v}$ is generated by rotating $\mathbf{u}$ into $\mathbf{v}$ and advancing in the direction of a right-handed screw. The magnitude of $\mathbf{u} \wedge \mathbf{v}$, given by

$$|\mathbf{u} \wedge \mathbf{v}| = uv \sin\varphi \tag{3.1.5.1}$$

is equal to the area of the parallelogram defined by $\mathbf{u}$ and $\mathbf{v}$.

It follows from the definition that

$$\mathbf{u} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{u}. \tag{3.1.5.2}$$

3.1.6. Permutation tensors

Many relationships involving vector products may be expressed compactly and conveniently in terms of the permutation tensors, defined as

$$\varepsilon_{ijk} = \mathbf{a}_i \cdot \mathbf{a}_j \wedge \mathbf{a}_k \tag{3.1.6.1}$$

$$\varepsilon^{ijk} = \mathbf{a}^i \cdot \mathbf{a}^j \wedge \mathbf{a}^k. \tag{3.1.6.2}$$

Since $\mathbf{a}_i \cdot \mathbf{a}_j \wedge \mathbf{a}_k$ represents the volume of the parallelepiped defined by vectors $\mathbf{a}_i$, $\mathbf{a}_j$, $\mathbf{a}_k$, it follows that ε_{ijk} vanishes if any two indices are equal to each other. The same argument applies, of course, to ε^{ijk}. That is,

$$\varepsilon_{ijk} = 0, \quad \varepsilon^{ijk} = 0, \quad \text{if } j = i \text{ or } k = i \text{ or } k = j. \tag{3.1.6.3}$$

If the indices are all different,

$$\varepsilon_{ijk} = PV, \quad \varepsilon^{ijk} = PV^* \tag{3.1.6.4}$$

for even permutations of *ijk* (123, 231, or 312), and

$$\varepsilon_{ijk} = -PV, \quad \varepsilon^{ijk} = -PV^* \tag{3.1.6.5}$$

for odd permutations (132, 213, or 321). Here, $P = +1$ for right-handed axes, $P = -1$ for left-handed axes, V is the unit-cell volume, and $V^* = 1/V$ is the volume of the reciprocal cell defined by the reciprocal basis vectors $\mathbf{a}^i$, $\mathbf{a}^j$, $\mathbf{a}^k$.

A discussion of the properties of the permutation tensors may be found in Sands (1982*a*). In right-handed Cartesian systems, where $P = 1$, and $V = V^* = 1$, the permutation tensors are equivalent to the permutation symbols denoted by e_{ijk}.

3.1.7. Components of vector product

As is shown in Sands (1982*a*), the components of the vector product $\mathbf{u} \wedge \mathbf{v}$ are given by

$$\mathbf{u} \wedge \mathbf{v} = \varepsilon_{ijk} u^i v^j \mathbf{a}^k, \tag{3.1.7.1}$$

where again $\mathbf{a}^k$ is a reciprocal basis vector (some writers use $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$ to represent the reciprocal axes). A special case of (3.1.7.1) is

$$\mathbf{a}_i \wedge \mathbf{a}_j = \varepsilon_{ijk} \mathbf{a}^k, \tag{3.1.7.2}$$

which may be taken as a defining equation for the reciprocal basis vectors. Similarly,

$$\mathbf{a}^i \wedge \mathbf{a}^j = \varepsilon^{ijk} \mathbf{a}_k, \tag{3.1.7.3}$$

which completes the characterization of the dual vector system with basis vectors $\mathbf{a}_i$ and $\mathbf{a}^j$ obeying

$$\mathbf{a}_i \cdot \mathbf{a}^j = \delta_i^j. \tag{3.1.7.4}$$

In (3.1.7.4), δ_i^j is the Kronecker delta, which equals 1 if $i = j$, 0 if $i \neq j$. The relationships between these quantities are explored at some length in Sands (1982*a*).

3.1.8. Some vector relationships

The results developed above lead to several useful relationships between vectors; for derivations, see Sands (1982*a*).

3.1.8.1. Triple vector product

$$\mathbf{u} \wedge (\mathbf{v} \wedge \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \tag{3.1.8.1}$$

$$(\mathbf{u} \wedge \mathbf{v}) \wedge \mathbf{w} = -(\mathbf{v} \cdot \mathbf{w})\mathbf{u} + (\mathbf{u} \cdot \mathbf{w})\mathbf{v}. \tag{3.1.8.2}$$

3.1.8.2. Scalar product of vector products

$$(\mathbf{u} \wedge \mathbf{v}) \cdot (\mathbf{w} \wedge \mathbf{z}) = (\mathbf{u} \cdot \mathbf{w})(\mathbf{v} \cdot \mathbf{z}) - (\mathbf{u} \cdot \mathbf{z})(\mathbf{v} \cdot \mathbf{w}). \quad (3.1.8.3)$$

A derivation of this result may be found also in Shmueli (1974).

3.1.8.3. Vector product of vector products

$$(\mathbf{u} \wedge \mathbf{v}) \wedge (\mathbf{w} \wedge \mathbf{z}) = (\mathbf{u} \cdot \mathbf{w} \wedge \mathbf{z})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w} \wedge \mathbf{z})\mathbf{u} \quad (3.1.8.4)$$

$$(\mathbf{u} \wedge \mathbf{v}) \wedge (\mathbf{w} \wedge \mathbf{z}) = (\mathbf{u} \cdot \mathbf{v} \wedge \mathbf{z})\mathbf{w} - (\mathbf{u} \cdot \mathbf{v} \wedge \mathbf{w})\mathbf{z}. \quad (3.1.8.5)$$

3.1.9. Planes

Among several ways of characterizing a plane in a general rectilinear coordinate system is a description in terms of the coordinates of three non-collinear points that lie in the plane. If the points are U, V and W, lying at the ends of vectors **u**, **v** and **w**, the vectors $\mathbf{u} - \mathbf{v}$, $\mathbf{v} - \mathbf{w}$ and $\mathbf{w} - \mathbf{u}$ are in the plane. The vector

$$\mathbf{z} = (\mathbf{u} - \mathbf{v}) \wedge (\mathbf{v} - \mathbf{w}) \quad (3.1.9.1)$$

is normal to the plane. Expansion of (3.1.9.1) yields

$$\mathbf{z} = (\mathbf{u} \wedge \mathbf{v}) + (\mathbf{v} \wedge \mathbf{w}) + (\mathbf{w} \wedge \mathbf{u}). \quad (3.1.9.2)$$

Making use of (3.1.7.1),

$$\mathbf{z} = \varepsilon_{ijk}(u^j v^k + v^j w^k + w^j u^k)\mathbf{a}^i. \quad (3.1.9.3)$$

If now **x** is any vector from the origin to the plane, $\mathbf{x} - \mathbf{u}$ is in the plane, and

$$(\mathbf{x} - \mathbf{u}) \cdot \mathbf{z} = 0. \quad (3.1.9.4)$$

From (3.1.9.2),

$$\mathbf{u} \cdot \mathbf{z} = \mathbf{u} \cdot \mathbf{v} \wedge \mathbf{w}. \quad (3.1.9.5)$$

Rearrangement of (3.1.9.4) with $\mathbf{x} \cdot \mathbf{z}$ on the left and $\mathbf{u} \cdot \mathbf{z}$ on the right, and using (3.1.9.3) for **z** on the left leads to

$$\varepsilon_{ijk}x^i(u^j v^k + v^j w^k + w^j u^k) = \varepsilon_{ijk}u^i v^j w^k. \quad (3.1.9.6)$$

If, in particular, the points are on the coordinate axes, their designations are $[u^1, 0, 0]$, $[0, v^2, 0]$ and $[0, 0, w^3]$, and (3.1.9.6) becomes

$$x^1/u^1 + x^2/v^2 + x^3/w^3 = 1, \quad (3.1.9.7)$$

which may be written

$$x^i h_i = 1 \quad (3.1.9.8)$$

or

$$\mathbf{x} \cdot \mathbf{h} = 1 \quad (3.1.9.9)$$

in which the vector **h** has coordinates

$$\mathbf{h} = (1/u^1, 1/v^2, 1/w^3). \quad (3.1.9.10)$$

That is, the covariant components of **h** are given by the reciprocals of the intercepts of the plane on the axes. The vector **h** is normal to the plane it describes (Sands, 1982*a*) and the length of **h** is the reciprocal of the distance d of the plane from the origin; *i.e.*,

$$h = 1/d. \quad (3.1.9.11)$$

If the indices h_i are relatively prime integers, the theory of numbers tells us that the Diophantine equation (3.1.9.8) has solutions x^i that are integers. Points whose contravariant components are integers are lattice points, and such a plane passes through an infinite number of lattice points and is called a *lattice plane*. Thus, the h_i for lattice planes are the familiar Miller indices of crystallography.

Calculations involving planes become quite manageable when the normal vector **h** is introduced. Thus, the distance l from a point P with coordinates p^i to a plane characterized by **h** is

$$l = (1 - \mathbf{p} \cdot \mathbf{h})/h, \quad (3.1.9.12)$$

where a negative sign indicates that the point is on the opposite side of the plane from the origin.

The dihedral angle τ between planes with normals **h** and $\mathbf{h}'$ is

$$\tau = \cos^{-1}[-\mathbf{h} \cdot \mathbf{h}'/(hh')]. \quad (3.1.9.13)$$

A variation of (3.1.9.13) expresses τ in terms of vector **u** in the first plane, vector **w** in the second plane, and vector **v**, the intersection of the planes, as (Shmueli, 1974)

$$\tau = \cos^{-1}[(\mathbf{u} \wedge \mathbf{v}) \cdot (\mathbf{v} \wedge \mathbf{w})/|\mathbf{u} \wedge \mathbf{v}||\mathbf{v} \wedge \mathbf{w}|]. \quad (3.1.9.14)$$

A similar calculation gives angles of torsion. Let $\mathbf{t}_h$ and $\mathbf{u}_h$ be, respectively, the projections of vectors **t** and **u** onto the plane with normal **h**.

$$\mathbf{t}_h = \mathbf{t} - (\mathbf{t} \cdot \mathbf{h})\mathbf{h}/h^2 \quad (3.1.9.15)$$

$$\mathbf{u}_h = \mathbf{u} - (\mathbf{u} \cdot \mathbf{h})\mathbf{h}/h^2. \quad (3.1.9.16)$$

The angle between $\mathbf{t}_h$ and $\mathbf{u}_h$ represents a torsion about **h** (Sands, 1982*b*). Another approach to the torsion angle, which gives equivalent results (Shmueli, 1974), is to compute the angle between $\mathbf{t} \wedge \mathbf{h}$ and $\mathbf{u} \wedge \mathbf{h}$ using (3.1.8.3).

3.1.10. Variance–covariance matrices

Refinement of a crystal structure yields both the parameters that describe the structure and estimates of the uncertainties of those parameters. Refinement by the method of least squares minimizes a weighted sum of squares of residuals. In the matrix notation of Hamilton's classic book (Hamilton, 1964), values of

the m parameters to be determined are expressed by the $m \times 1$ column vector $\boldsymbol{X}$ given by

$$\boldsymbol{X} = (\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{F}, \tag{3.1.10.1}$$

where $\boldsymbol{F}$ is an $n \times 1$ matrix representing the observations (structure factors or squares of structure factors), $\boldsymbol{P}$ is an $n \times n$ weight matrix that is proportional to the variance–covariance matrix of the observed $\boldsymbol{F}$, $\boldsymbol{A}$ is an $n \times m$ design matrix consisting of the derivatives of each element of $\boldsymbol{F}$ with respect to each of the parameters and $\boldsymbol{A}^T$ is the transpose of $\boldsymbol{A}$. The variance–covariance matrix of the parameters is then given by

$$\boldsymbol{M} = \boldsymbol{V}^T\boldsymbol{P}\boldsymbol{V}(\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A})^{-1}/(n-m). \tag{3.1.10.2}$$

Here, $\boldsymbol{V}$ is the $n \times 1$ matrix of residuals, consisting of the differences between the observed and calculated values of the elements of $\boldsymbol{F}$. Since $\boldsymbol{V}^T\boldsymbol{P}\boldsymbol{V}/(n-m)$ is just a single number, $\boldsymbol{M}$ is proportional to the inverse least-squares matrix $(\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A})^{-1}$.

Once the variance–covariance matrix of the parameters is known, the variances and covariances of any quantities derived from these parameters can be computed. The variance of a single function f is given by

$$\sigma^2(f) = \frac{\partial f}{\partial x^i}\frac{\partial f}{\partial x^j}\mathrm{cov}(x^i, x^j), \tag{3.1.10.3}$$

where, as usual, we are using the summation convention and summing over all parameters included in f. A generalization of (3.1.10.3) for two functions is

$$\mathrm{cov}(f_1, f_2) = \frac{\partial f_1}{\partial x^i}\frac{\partial f_2}{\partial x^j}\mathrm{cov}(x^i, x^j). \tag{3.1.10.4}$$

[The covariance of two quantities is, of course, just the variance if the two quantities are the same. For an elementary discussion of statistical covariance and correlation, see Sands (1977).] Equation (3.1.10.4) may now be extended to any number of functions (Sands, 1966); the $k \times k$ variance–covariance matrix $\boldsymbol{C}$ of k functions of m parameters is given in terms of the $m \times m$ variance–covariance matrix of the parameters by

$$\boldsymbol{C} = \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}^T, \tag{3.1.10.5}$$

in which the ijth element of the $k \times m$ matrix $\boldsymbol{D}$ is the derivative of function f_i with respect to parameter j. Element C_{II} (no summation implied over I) is the variance of function f_I, and C_{IJ} is the covariance of functions f_I and f_J.

The calculation of $\boldsymbol{C}$ must, of course, include the contributions of all sources of error, so $\boldsymbol{M}$ in (3.1.10.5) should include the variances and covariances of the unit-cell dimensions and of any other relevant parameters with non-negligible uncertainties.

It may be easier, in some cases, to carry out calculations of variances and covariances in steps. For example, the variance–covariance matrix of a set of distances may be computed and then other quantities may be determined as functions of the distances. It is imperative that all nonvanishing covariances be included in every stage of the calculation; only in special cases are the covariances negligible, and often they are large enough to affect the results seriously (Sands, 1977).

These principles may be used to explore the effects of symmetry or of transformations on the variance–covariance matrices of atomic parameters and derived quantities. Using the notation of Sands (1966), with x_A^i and x_B^i the positional parameters i of atoms A and B, respectively, we define $\boldsymbol{M}_{AA}$, $\boldsymbol{M}_{AB}$, $\boldsymbol{M}_{BA}$ and $\boldsymbol{M}_{BB}$ as 3×3 matrices with ijth elements $\mathrm{cov}(x_A^i, x_A^j)$, $\mathrm{cov}(x_A^i, x_B^j)$, $\mathrm{cov}(x_B^i, x_A^j)$ and $\mathrm{cov}(x_B^i, x_B^j)$, respectively. If atom B' is generated from atom B by symmetry operator $\boldsymbol{S}$, such that

$$\mathbf{x}_{B'} = \boldsymbol{S}\mathbf{x}_B \tag{3.1.10.6}$$

$$x_{B'}^i = S_j^i x_B^j, \tag{3.1.10.7}$$

it is shown in Sands (1966) that the variance–covariance matrices involving atom B' are

$$\boldsymbol{M}_{AB'} = \boldsymbol{M}_{AB}\boldsymbol{S}^T \tag{3.1.10.8}$$

$$\boldsymbol{M}_{B'A} = \boldsymbol{S}\boldsymbol{M}_{BA} \tag{3.1.10.9}$$

$$\boldsymbol{M}_{B'B'} = \boldsymbol{S}\boldsymbol{M}_{BB}\boldsymbol{S}^T. \tag{3.1.10.10}$$

If symmetry operator $\boldsymbol{S}$ is applied to both atoms A and B to generate atoms A' and B', the corresponding matrices may be expressed by the matrix equation

$$\begin{pmatrix} \boldsymbol{M}_{A'A'} & \boldsymbol{M}_{A'B'} \\ \boldsymbol{M}_{B'A'} & \boldsymbol{M}_{B'B'} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}\boldsymbol{M}_{AA}\boldsymbol{S}^T & \boldsymbol{S}\boldsymbol{M}_{AB}\boldsymbol{S}^T \\ \boldsymbol{S}\boldsymbol{M}_{BA}\boldsymbol{S}^T & \boldsymbol{S}\boldsymbol{M}_{BB}\boldsymbol{S}^T \end{pmatrix}. \tag{3.1.10.11}$$

If $\boldsymbol{G}$ is a matrix that transforms to a new set of axes,

$$\mathbf{a}' = \boldsymbol{G}\mathbf{a}, \tag{3.1.10.12}$$

the transformed variance–covariance matrix of the atomic parameters is

$$\boldsymbol{M}' = (\boldsymbol{G}^{\boldsymbol{T}})^{-1}\boldsymbol{M}\boldsymbol{G}^{-1}. \tag{3.1.10.13}$$

To apply these formulae to calculations of the errors and covariances of interatomic distances and angles, consider the triangle of atoms A, B, C with edges $l_1 = AB$, $l_2 = BC$, $l_3 = CA$, and angles $\alpha_1, \alpha_2, \alpha_3$ at A, B, C, respectively. If the atoms are not related by symmetry,

$$\sigma^2(l_1) = \boldsymbol{l}_1^T\mathbf{g}(\boldsymbol{M}_{AA} - \boldsymbol{M}_{AB} - \boldsymbol{M}_{BA} + \boldsymbol{M}_{BB})\mathbf{g}\boldsymbol{l}_1/l_1^2 \tag{3.1.10.14}$$

$$\mathrm{cov}(l_1, l_2) = \boldsymbol{l}_1^T\mathbf{g}(\boldsymbol{M}_{AB} - \boldsymbol{M}_{AC} - \boldsymbol{M}_{BB} + \boldsymbol{M}_{BC})\mathbf{g}\boldsymbol{l}_2/l_1 l_2. \tag{3.1.10.15}$$

If atom B is generated from atom A by symmetry matrix $\boldsymbol{S}$, the results, as derived in Sands (1966), are

$$\sigma^2(l_1) = \boldsymbol{l}_1^T\mathbf{g}(\boldsymbol{M}_{AA} - \boldsymbol{S}\boldsymbol{M}_{AA} - \boldsymbol{M}_{AA}\boldsymbol{S}^T + \boldsymbol{S}\boldsymbol{M}_{AA}\boldsymbol{S}^T)\mathbf{g}\boldsymbol{l}_1/l_1^2 \quad (3.1.10.16)$$

$$\sigma^2(l_2) = \boldsymbol{l}_2^T\mathbf{g}(\boldsymbol{S}\boldsymbol{M}_{AA}\boldsymbol{S}^T - \boldsymbol{M}_{AC}\boldsymbol{S}^T - \boldsymbol{S}\boldsymbol{M}_{AC} + \boldsymbol{M}_{CC})\mathbf{g}\boldsymbol{l}_2/l_2^2 \quad (3.1.10.17)$$

$$\sigma^2(l_3) = \boldsymbol{l}_3^T\mathbf{g}(\boldsymbol{M}_{AA} - \boldsymbol{M}_{AC} - \boldsymbol{M}_{CA} + \boldsymbol{M}_{CC})\mathbf{g}\boldsymbol{l}_3/l_3^2 \quad (3.1.10.18)$$

$$\mathrm{cov}(l_1, l_2) = \boldsymbol{l}_1^T\mathbf{g}(\boldsymbol{M}_{AA}\boldsymbol{S}^T - \boldsymbol{S}\boldsymbol{M}_{AA}\boldsymbol{S}^T - \boldsymbol{M}_{AC} + \boldsymbol{S}\boldsymbol{M}_{AC})\mathbf{g}\boldsymbol{l}_2/l_1l_2 \quad (3.1.10.19)$$

$$\mathrm{cov}(l_1, l_3) = \boldsymbol{l}_1^T\mathbf{g}(-\boldsymbol{M}_{AA} + \boldsymbol{S}\boldsymbol{M}_{AA} + \boldsymbol{M}_{AC} - \boldsymbol{S}\boldsymbol{M}_{AC})\mathbf{g}\boldsymbol{l}_3/l_1l_3 \quad (3.1.10.20)$$

$$\mathrm{cov}(l_2, l_3) = \boldsymbol{l}_2^T\mathbf{g}(-\boldsymbol{S}\boldsymbol{M}_{AA} + \boldsymbol{M}_{CA} + \boldsymbol{S}\boldsymbol{M}_{AC} - \boldsymbol{M}_{CC})\mathbf{g}\boldsymbol{l}_3/l_2l_3. \quad (3.1.10.21)$$

In equations (3.1.10.14)–(3.1.10.21), $\boldsymbol{l}_i$ is a column vector with components the differences of the coordinates of the atoms connected by the vector. Representative formulae involving the angles α_1, α_2, α_3 are

$$\sigma^2(\alpha_1) = [\cos^2\alpha_2\sigma^2(l_1) - 2\cos\alpha_2\ \mathrm{cov}(l_1, l_2) + 2\cos\alpha_2\cos\alpha_3\ \mathrm{cov}(l_1, l_3) + \sigma^2(l_2) - 2\cos\alpha_3\ \mathrm{cov}(l_2, l_3) + \cos^2\alpha_3\sigma^2(l_3)](l_2/l_1l_3\sin\alpha_1)^2 \quad (3.1.10.22)$$

$$\mathrm{cov}(\alpha_1, \alpha_2) = [\cos\alpha_1\cos\alpha_2\sigma^2(l_1) + (\cos\alpha_2\cos\alpha_3 - \cos\alpha_1)\ \mathrm{cov}(l_1, l_2) + (\cos\alpha_1\cos\alpha_3 - \cos\alpha_2)\ \mathrm{cov}(l_1, l_3) - \cos\alpha_3\sigma^2(l_2) + (1 + \cos^2\alpha_3)\ \mathrm{cov}(l_2, l_3) - \cos\alpha_3\sigma^2(l_3)]/(l_1^2\sin\alpha_1\sin\alpha_2) \quad (3.1.10.23)$$

$$\mathrm{cov}(\alpha_1, l_1) = [-\cos\alpha_2\sigma^2(l_1) + \mathrm{cov}(l_1, l_2) - \cos\alpha_3\ \mathrm{cov}(l_1, l_3)](l_2/l_1l_3\sin\alpha_1) \quad (3.1.10.24)$$

$$\mathrm{cov}(\alpha_1, l_2) = [-\cos\alpha_2\ \mathrm{cov}(l_1, l_2) + \sigma^2(l_2) - \cos\alpha_3\ \mathrm{cov}(l_2, l_3)](l_2/l_1l_3\sin\alpha_1). \quad (3.1.10.25)$$

If any of the angles approach 0 or 180°, the denominators in (3.1.10.22)–(3.1.10.25) will become very small, necessitating high-precision arithmetic. Indeterminacies resulting from special relationships between atomic positions may require rederivation of the equations for variances and covariances, to take the relationships into account explicitly and avoid the indeterminacies. A true symmetry condition requiring, for example, a linear bond should cause little problem, and the corresponding variance will be zero. It is the indeterminacies not originating from crystal symmetry that demand caution, in recognizing them and in coping with them correctly.

A general expression for the variance of a dihedral angle, in terms of the variances and covariances of the coordinates of the four atoms, is (Shmueli, 1974)

$$\sigma^2(\tau) = \sum_k\sum_n \frac{\partial\tau}{\partial x^i_{(k)}}\frac{\partial\tau}{\partial x^j_{(n)}}\mathrm{cov}[x^i_{(k)}, x^j_{(n)}], \quad (3.1.10.26)$$

where, in addition to the usual tensor summation over i and j from 1 to 3, summation must be carried out over the four atoms (*i.e.*, k and n vary from 1 to 4). Special cases of (3.1.10.26), corresponding to various levels of approximation of diagonal matrices and isotropic errors, are given in Shmueli (1974). Formulae in dyadic notation are given in Waser (1973) for the variances and covariances of dihedral angles, of best planes, of torsion angles, and of other molecular parameters.

3.1.11. Mean values

The weighted mean of a set of quantities X_i is

$$\langle X\rangle = \sum w_iX_i/\sum w_i, \quad (3.1.11.1)$$

where the weights are typically chosen to minimize the variance of $\langle X\rangle$. The variance may be computed from the variance–covariance matrix $\boldsymbol{M}$ of the X_i by

$$\sigma^2(\langle X\rangle) = \mathbf{w}^T\boldsymbol{M}\mathbf{w}/(\sum w_i)^2. \quad (3.1.11.2)$$

Minimization of $\sigma^2(\langle X\rangle)$ leads to weights given by

$$\mathbf{w} = \boldsymbol{M}^{-1}\mathbf{v}, \quad (3.1.11.3)$$

where the components of vector $\mathbf{v}$ are all equal ($v_i = v_j$ for all i and j); since (3.1.11.1) and (3.1.11.2) require only relative weights, we can assign $v_i = 1$ for all i. Placing these weights in (3.1.11.2) yields

$$\sigma^2(\langle X\rangle) = 1/\sum w_i. \quad (3.1.11.4)$$

For the case of uncorrelated X_i, the weights are inversely proportional to the corresponding variances

$$w_i = 1/\sigma^2(X_i). \quad (3.1.11.5)$$

For the case of two correlated variables,

$$w_i = 1/[\sigma^2(X_i) - \mathrm{cov}(X_1, X_2)]. \quad (3.1.11.6)$$

Derivation and discussion of these equations may be found in Sands (1966, 1982*b*).

The presence of systematic errors in the experimental data often results in these formulae producing estimates of the standard uncertainties of molecular dimensions that are too small; it has been suggested that such error estimates should be multiplied by 1.5 to make them more realistic (Taylor & Kennard, 1983). It is essential also that averages be computed only of similar quantities, and interatomic distances corresponding to different bond orders or different environments may not represent the same physical quantities; that is, there are reasons for the discrepancies, and averaging may obscure important information. Another source of error in molecular geometry parameters determined from crystallographic measurements is thermal motion, and distances should be corrected for such effects before making comparisons (Busing & Levy, 1964; Johnson, 1970, 1980).

A discussion of the appropriateness of weighted and unweighted means may be found in Taylor & Kennard (1985), which suggests that the unweighted mean might even be preferable if environmental effects are large.

3.1.12. Computation

It has been mentioned that the tensor formulation used in this chapter is particularly amenable to machine computation. As a simple illustration of this point, the following Fortran program will compute the lengths of vectors **X** and **Y** and the angle between them.

```
      DIMENSION X(3),Y(3),G(3,3),SUM(3)
      READ (5,10)(X(I),I = 1,3)
      READ (5,10)(Y(I),I = 1,3)
      READ (5,10)((G(I,J),J = 1,3),I = 1,3)
10    FORMAT (3F10.5)
      DO 20 I = 1,3
20    SUM(I) = 0
      DO 30 I = 1,3
      DO 30 J = 1,3
      SUM(1) = SUM(1) + X(I) * X(J) * G(I,J)
      SUM(2) = SUM(2) + Y(I) * Y(J) * G(I,J)
      SUM(3) = SUM(3) + X(I) * Y(J) * G(I,J)
30    CONTINUE
      DIST1 = SQRT(SUM(1))
      DIST2 = SQRT(SUM(2))
      ANGLE = 57.296 * ACOS(SUM(3)/(DIST1 * DIST2))
      WRITE (6,10) DIST1,DIST2,ANGLE
      END
```

References

Busing, W. R. & Levy, H. A. (1964). *Effect of thermal motion on the estimation of bond lengths. Acta Cryst.* **17**, 142–146.

Hamilton, W. C. (1964). *Statistics in Physical Science*. New York: Ronald Press.

Johnson, C. K. (1970). *The effect of thermal motion on interatomic distances and angles*. In *Crystallographic Computing*, edited by F. R. Ahmed, pp. 220–226. Copenhagen: Munksgaard.

Johnson, C. K. (1980). *Thermal motion analysis*. In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 14.01–14.19. Bangalore: Indian Academy of Sciences.

Sands, D. E. (1966). *Transformations of variance–covariance tensors. Acta Cryst.* **21**, 868–872.

Sands, D. E. (1977). *Correlation and covariance. J. Chem. Educ.* **54**, 90–94.

Sands, D. E. (1982*a*). *Vectors and Tensors in Crystallography*. Reading: Addison Wesley. Reprinted (1995) Dover Publications.

Sands, D. E. (1982*b*). *Molecular geometry*. In *Computational Crystallography*, edited by D. Sayre, pp. 421–429. Oxford: Clarendon Press.

Shmueli, U. (1974). *On the standard deviation of a dihedral angle. Acta Cryst.* A**30**, 848–849.

Taylor, R. & Kennard, O. (1983). *The estimation of average molecular dimensions from crystallographic data. Acta Cryst.* B**39**, 517–525.

Taylor, R. & Kennard, O. (1985). *The estimation of average molecular dimensions. 2. Hypothesis testing with weighted and unweighted means. Acta Cryst.* A**41**, 85–89.

Waser, J. (1973). *Dyadics and the variances and covariances of molecular parameters, including those of best planes. Acta Cryst.* A**29**, 621–631.

International Tables for Crystallography (2010). Vol. B, Chapter 3.2, pp. 410–417.

3.2. The least-squares plane

By R. E. Marsh and V. Schomaker†

3.2.1. Introduction

By way of introduction, we remark that in earlier days of crystal structure analysis, before the advent of high-speed computers and routine three-dimensional analyses, molecular planarity was often assumed so that atom coordinates along the direction of projection could be estimated from two-dimensional data [see, *e.g.*, Robertson (1948)]. Today, the usual aim in deriving the coefficients of a plane is to investigate the degree of planarity of a group of atoms as found in a full, three-dimensional structure determination. We further note that, for such purposes, a crystallographer will often be served just as well by establishing the plane in an almost arbitrary fashion as by resorting to the most elaborate, nit-picking and pretentious least-squares treatment. The approximate plane and the associated perpendicular distances of the atoms from it will be all he needs as scaffolding for his geometrical and structural imagination; reasonable common sense will take the place of explicit attention to error estimates.

Nevertheless, we think it appropriate to lay out in some detail the derivation of the 'best' plane, in a least-squares sense, through a group of atoms and of the standard uncertainties associated with this plane. We see two cases: (1) The weights of the atoms in question are considered to be isotropic and uncorrelated (*i.e.* the weight matrix for the positions of all the atoms is diagonal, when written in terms of Cartesian axes, and for each atom the three diagonal elements are equal). In such cases the weights may have little or nothing to do with estimates of random error in the atom positions (they may have been assigned merely for convenience or convention), and, therefore, no one should feel that the treatment is proper in respect to the theory of errors. Nevertheless, it may be desired to incorporate the error estimates (variances) of the atom positions into the *results* of such calculations, whereupon these variances (which may be anisotropic, with correlation between atoms) need to be propagated. In this case the distinction between *weights* (or their inverses) and *variances* must be kept very clear. (2) The weights are anisotropic and are presumably derived from a variance–covariance matrix, which may include correlation terms between different atoms; the objective is to achieve a truly proper Gaussian least-squares result.

3.2.2. Least-squares plane based on uncorrelated, isotropic weights

This is surely the most common situation; it is not often that one will wish to take the trouble, or be presumptive enough, to assign anisotropic or correlated weights to the various atoms. And one will sometimes, perhaps even often, not be genuinely interested in the hypothesis that the atoms actually are rigorously coplanar; for instance, one might be interested in examining the best plane through such a patently nonplanar molecule as cyclohexane. Moreover, the calculation is simple enough, given the availability of computers and programs, as to be a practical realization of the off-the-cuff treatment suggested in our opening paragraph. The problem of deriving the plane's coefficients is intrinsically nonlinear in the way first discussed by Schomaker *et al.* (1959; SWMB). Any formulation other than as an eigenvalue–eigenvector problem (SWMB), as far as we can tell, will sometimes go astray. As to the propagation of errors, numerous treatments have been given, but none that we have seen is altogether satisfactory.

We refer all vectors and matrices to Cartesian axes, because that is the most convenient in calculation. However, a more elegant formulation can be written in terms of general axes [*e.g.*, as in Shmueli (1981)].

The notation is troublesome. Indices are needed for atom number and Cartesian direction, and the exponent 2 is needed as well, which is difficult if there are superscript indices. The best way seems to be to write all the indices as subscripts and distinguish among them by context – $i, j, 1, 2, 3$ for directions; k, l, p (and sometimes $K, \ldots$) for atoms. In any case, *atom* first then *direction* if there are two subscripts; *direction*, if only one index for a vector component, but *atom* (in this section at least) if for a weight or a vector. And σ_{d_1}, *e.g.*, for the standard uncertainty of the distance of atom 1 from a plane. For simplicity in practice, we use Cartesian coordinates throughout.

The first task is to find the plane, which we write as

$$0 = \mathbf{m} \cdot \mathbf{r} - d \equiv \mathsf{m}^T\mathsf{r} - d,$$

where $\mathbf{r}$ is here the vector from the origin to any point on the plane (but usually represents the measured position of an atom), $\mathbf{m}$ is a unit vector parallel to the normal from the origin to the plane, d is the length of the normal, and m and r are the column representations of $\mathbf{m}$ and $\mathbf{r}$. The least-squares condition is to find the stationary values of $S \equiv [w_k(\mathsf{m}^T\mathsf{r}_k - d)^2]$ subject to $\mathsf{m}^T\mathsf{m} = 1$, with r_k, $k = 1, \ldots, n$, the vector from the origin to atom k and with weights, w_k, isotropic and without interatomic correlations for the n atoms of the plane. We also write S as $S \equiv [w(\mathsf{m}^T\mathsf{r} - d)^2]$, the subscript for atom number being implicit in the Gaussian summations ($[\ldots]$) over all atoms, as it is also in the angle-bracket notation for the weighted average over all atoms, for example in $\langle\mathsf{r}\rangle$ – the weighted centroid of the groups of atoms – just below.

First solve for d, the origin-to-plane distance.

$$0 = -\frac{1}{2}\frac{\partial S}{\partial d} = [w(\mathsf{m}^T\mathsf{r} - d)] = 0,$$
$$d = [w\mathsf{m}^T\mathsf{r}]/[w] \equiv \mathsf{m}^T\langle\mathsf{r}\rangle.$$

Then

$$S \equiv [w(\mathsf{m}^T\mathsf{r} - d)^2] = [w\{\mathsf{m}^T(\mathsf{r} - \langle\mathsf{r}\rangle)\}^2]$$
$$\equiv [w(\mathsf{m}^T\mathbf{s})^2] \equiv \mathsf{m}^T[w\mathsf{s}\mathsf{s}^T]\mathsf{m} \equiv \mathsf{m}^T\mathbf{A}\mathsf{m}.$$

Here $\mathbf{s}_k \equiv \mathsf{r}_k - \langle\mathsf{r}\rangle$ is the vector from the centroid to atom k. Then solve for $\mathbf{m}$. This is the eigenvalue problem – to diagonalize $\mathbf{A}$ (bear in mind that $\mathbf{A}_{ij}$ is just $[ws_is_j]$) by rotating the coordinate axes, *i.e.*, to find the 3×3 arrays $\mathbf{M}$ and $\mathbf{L}$, $\mathbf{L}$ diagonal, to satisfy

$$\mathbf{M}^T\mathbf{A}\mathbf{M} = \mathbf{L}, \qquad \mathbf{M}^T\mathbf{M} = \mathbf{I}.$$

$\mathbf{A}$ and $\mathbf{M}$ are symmetric; the columns m of $\mathbf{M}$ are the direction cosines of, and the diagonal elements of $\mathbf{L}$ are the sums of

† Deceased.

Copyright © 2010 International Union of Crystallography

weighted squares of residuals from, the best, worst and intermediate planes, as discussed by SWMB.

3.2.2.1. Error propagation

Waser *et al.* (1973; WMC) carefully discussed how the random errors of measurement of the atom positions propagate into the derived quantities in the foregoing determination of a least-squares plane. This section presents an extension of their discussion. To begin, however, we first show how standard first-order perturbation theory conveniently describes the propagation of error into **M** and **L** when the positions r_k of the atoms are incremented by the amounts $\delta\mathsf{r}_k \equiv \xi_k$ and the corresponding quantities $\mathbf{s}_k \equiv \mathsf{r}_k - \langle\mathsf{r}\rangle$ (the vectors from the centroid to the atoms) by the amounts η_k, $(\mathbf{s} \rightarrow \mathbf{s} + \eta)$, $\eta_k \equiv \xi_k - \langle\xi\rangle$. (The need to account for the variation in position of the centroid, *i.e.* to distinguish between η_k and ξ_k, was overlooked by WMC.) The consequent increments in **A**, **M** and **L** are

$$\begin{aligned}\delta\mathbf{A} &= [w\eta\mathbf{s}^T] + [w\mathbf{s}\eta^T] \equiv \alpha,\\ \delta\mathbf{M} &= \mathbf{M}\mu,\\ \delta\mathbf{L} &\equiv \lambda.\end{aligned}$$

Here the columns of μ are expressed as linear combinations of the columns of **M**. Note also that both perturbations, μ and λ, which are the adjustments to the orientations and associated eigenvalues of the principal planes, will depend on the reduced coordinates **s** and the perturbing influences ξ by way of α, which in turn depends only on the reduced coordinates and the reduced shifts η_k. In contrast,

$$\delta\mathsf{d} = \delta(\mathsf{m}^T\langle\mathsf{r}\rangle) = (\delta\mathsf{m}^T)\langle\mathsf{r}\rangle + \mathsf{m}^T\langle\xi\rangle;$$

the change in the origin-to-plane distance for the plane defined by the column vectors **m** of **M**, depends on the $\langle\mathsf{r}\rangle$ and $\langle\xi\rangle$ directly as well as on the **s** and η by way of the $\delta\mathbf{m}$.

The first-order results arising from the standard conditions, $\mathbf{M}^T\mathbf{M} = \mathbf{I}$, **L** diagonal, and $\mathbf{M}^T\mathbf{A}\mathbf{M} = \mathbf{L}$, are

$$\mu^T + \mu = 0, \lambda \text{ diagonal},$$

and

$$\mu^T\mathbf{M}^T\mathbf{A}\mathbf{M} + \mathbf{M}^T\alpha\mathbf{M} + \mathbf{M}^T\mathbf{A}\mathbf{M}\mu = \mu^T\mathbf{L} + \mathbf{L}\mu + \mathbf{M}^T\alpha\mathbf{M} = \lambda.$$

Stated in terms of the matrix components λ_{ij} and μ_{ij}, the first condition is $\mu_{ij} = -\mu_{ji}$, hence $\mu_{ii} = 0$, $i, j = 1, 2, 3$, and the second is $\lambda_{ij} = 0$, $i \neq j$. With these results the third condition then reads

$$\begin{aligned}\lambda_{jj} &= (\mathbf{M}^T\alpha\mathbf{M})_{jj}, && j = 1, 2, 3\\ \mu_{ij} &= (\mathbf{M}^T\alpha\mathbf{M})_{ij}/(L_{jj} - L_{ii}), && i \neq j, \quad i, j = 1, 2, 3.\end{aligned}$$

All this is analogous to the usual first-order perturbation theory, as, for example, in elementary quantum mechanics.

Now rotate to the coordinates defined by WMC, with axes parallel to the original eigenvectors [$\mathbf{M} = \mathbf{I}$, $A_{ij} = L_{ij}\delta_{ij}$, $(\mathbf{M}^T\alpha\mathbf{M})_{ij} = \alpha_{ij}$], restrict attention to the best plane ($M_{13} \equiv m_1 = 0$, $M_{23} \equiv m_2 = 0$, $M_{33} \equiv m_3 = 1$), and define ε^T as $(\delta m_1, \delta m_2, \delta d_c)$, keeping in mind $\delta m_3 = \mu_{33} = 0$; d_c itself, the original plane-to-centroid distance, of course vanishes. One then finds

$$\begin{aligned}\delta m_i \equiv \varepsilon_i &= \alpha_{i3}/(L_{33} - L_{ii})\\ &= [w(s_i\eta_3 + s_3\eta_i)]/([ws_3^2] - [ws_i^2]), \quad i = 1, 2,\\ \delta d_c \equiv \varepsilon_3 &= [w\xi_3]/[w] \equiv \langle\xi_3\rangle, \qquad \xi_k \equiv \delta r_k,\end{aligned}$$

and also

$$\delta d = \varepsilon_1\langle r_1\rangle + \varepsilon_2\langle r_2\rangle + \varepsilon_3.$$

These results have simple interpretations. The changes in direction of the plane normal (the δm_i) are rotations, described by ε_1 and ε_2, in response to changes in moments acting against effective torsion force constants. For ε_2, for example, the contribution of atom k to the total relevant moment, about direction 1, is $-w_k s_{k3} s_{k2}$ ($w_k s_{k3}$ the 'force' and s_{k2} the lever arm), and its nominally first-order change has two parts, $-w_k s_{k2}\eta_3$ from the change in force and $-w_k s_{k3}\eta_2$ from the change in lever arm; the resisting torsion constant is $[ws_2^2] - [ws_3^2]$, which, reflection will show, is qualitatively reasonable if not quantitatively obvious. The perpendicular displacement of the plane from the original centroid $\langle r\rangle$ is ε_3, but there are two further contributions to δd, the change in distance from origin to plane along the plane normal, that arise from the two components of out-of-plane rotation of the plane about its centroid. Note that ε_3 is not given by $[w\eta_3]/[w] = [w(\xi_3 - \langle\xi_3\rangle)]/[w]$, which vanishes identically.

There is a further, somewhat delicate point: If the group of atoms is indeed essentially coplanar, the s_{k3} are of the same order of magnitude as the η_{ki}, unlike the s_{ki}, $i \neq 3$, which are in general about as big as the lateral extent of the group. It is then appropriate to drop all terms in η_i or ξ_i, $i \neq 3$, and, in the denominators, the terms in s_{k3}^2.

The covariances of the derived quantities (by *covariances* we mean here both variances and covariances) can now be written out rather compactly by extending the implicit designation of atom numbers to double sums, the first of each of two similar factors referring to the first atom index and the second to the second, *e.g.*, $\sum_{kl} ww(s_i s_j)\ldots \equiv \sum_{kl} w_k w_l (s_{ki} s_{ij})\ldots$. Note that the various covariances, *i.e.* the averages over the presumed population of random errors of replicated measurements, are indicated by overlines, angle brackets having been pre-empted for averages over sets of atoms.

$$\begin{aligned}\operatorname{cov}(m_i, m_j) &\equiv \overline{\varepsilon_i\varepsilon_j}\\ &= \frac{\sum_{kl} ww(s_i s_j\overline{\eta_3\eta_3} + s_3 s_3\overline{\eta_i\eta_j} + s_i s_3\overline{\eta_3\eta_j} + s_3 s_j\overline{\eta_i\eta_3})}{\{[w(s_3^2 - s_i^2)]\}\{[w(s_3^2 - s_j^2)]\}}, \; i, j = 1, 2\\ \operatorname{cov}(m_i, d_c) \equiv \overline{\varepsilon_i\varepsilon_3} &= \frac{\sum_{kl} ww(s_{ki}\overline{\eta_3\xi_3} + s_{k3}\overline{\eta_i\xi_3})}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i, j = 1, 2\\ \sigma^2(d_c) \equiv \overline{\varepsilon_3^2} &= \frac{\sum_{kl} ww\overline{\xi_3\xi_3}}{[w]^2}\\ \sigma^2(d) \equiv \langle(\delta d)^2\rangle &= \langle r_1\rangle^2\overline{\varepsilon_1^2} + \langle r_2\rangle^2\overline{\varepsilon_2^2} + \overline{\varepsilon_3^2} + 2\langle r_1\rangle\langle r_2\rangle\overline{\varepsilon_1\varepsilon_2}\\ &\quad + 2\langle r_1\rangle\overline{\varepsilon_1\varepsilon_3} + 2\langle r_2\rangle\overline{\varepsilon_2\varepsilon_3}.\end{aligned}$$

Interatomic covariance (*e.g.*, $\overline{\eta_{k3}\eta_{l3}}$, $k \neq l$) thus presents no formal difficulty, although actual computation may be tedious. Nonzero covariance for the η's *may* arise explicitly from interatomic covariance (*e.g.*, $\overline{\xi_{ki}\xi_{lj}}$, $k \neq l$) of the errors in the atomic positions r_k, and it will *always* arise implicitly because $\langle\xi\rangle$ in $\eta_k = \xi_k - \langle\xi\rangle$ includes all the ξ_k and therefore has nonzero covariance with all of them and with itself, even if there is no interatomic covariance among the ξ_i's.

If both types of interatomic covariance (explicit and implicit) are negligible, the ε covariances simplify a great deal, the double

summations reducing to single summations. [The formal expression for $\sigma^2(d)$ does not change, so it will not be repeated.]

$$\begin{aligned}
\mathrm{cov}(m_i, m_j) &\equiv \overline{\varepsilon_i \varepsilon_j} \\
&= \frac{[w^2(s_i s_j \overline{\eta_3^2} + s_3^2 \overline{\eta_i \eta_j} + s_i s_3 \overline{\eta_3 \eta_j} + s_3 s_j \overline{\eta_i \eta_3}]}{\{[w(s_3^2 - s_i^2)]\}\{[w(s_3^2 - s_j^2)]\}}, \; i, j = 1, 2 \\
\mathrm{cov}(m_i, d_c) &\equiv \overline{\varepsilon_i \varepsilon_3} = \frac{[w^2(s_i \overline{\eta_3 \xi_3} + s_3 \overline{\eta_i \xi_3})]}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i, j = 1, 2 \\
\sigma^2(d_c) &\equiv \overline{\varepsilon_3^2} = \frac{[w^2 \overline{\xi_3^2}]}{[w]^2}.
\end{aligned}$$

When the variances are the same for ξ as for η (*i.e.* $\overline{\xi_i \xi_j} = \overline{\eta_i \eta_j}$, all i, j) and the covariances all vanish ($\overline{\xi_i \xi_j} = 0, \; i \neq j$), the $\overline{\varepsilon_i \varepsilon_j}$ simplify further. If the variances are also isotropic ($\overline{\xi_i^2} = \overline{\xi_j^2} = \sigma^2$, all i, j), there is still further simplification to

$$\begin{aligned}
\sigma^2(m_i) \equiv \overline{\varepsilon_i^2} &= \frac{[w^2 \sigma^2 (s_i^2 + s_3^2)]}{\{[w(s_3^2 - s_i^2)]\}^2}, \quad i = 1, 2 \\
\sigma^2(d_c) \equiv \overline{\varepsilon_3^2} &= [w^2 \sigma^2]/[w]^2 \\
\mathrm{cov}(m_1, m_2) \equiv \overline{\varepsilon_1 \varepsilon_2} &= \frac{[w^2 \sigma^2 s_1 s_2]}{\{[w(s_3^2 - s_1^2)]\}\{[w(s_3^2 - s_2^2)]\}} \\
\mathrm{cov}(m_i, d_c) \equiv \overline{\varepsilon_i \varepsilon_3} &= \frac{[w^2 \sigma^2 s_i]}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i = 1, 2.
\end{aligned}$$

If allowance is made for the difference in definition between ε_3 and δd, these expressions are equivalent to the ones (equations 7–9) given by WMC, who, however, do not appear to have been aware of the distinction between η and ξ and the possible consequences thereof.

If, finally, w^{-1} for each atom is taken equal to its $\overline{\eta_j^2} = \sigma^2$, all j, there is still further simplification.

$$\begin{aligned}
\sigma^2(m_i) \equiv \overline{\varepsilon_i^2} &= \frac{[w(s_3^2 + s_i^2)]}{\{[w(s_3^2 - s_i^2)]\}^2}, \quad i = 1, 2 \\
\sigma^2(d_c) \equiv \overline{\varepsilon_3^2} &= [w]/[w]^2 = 1/[w] \\
\mathrm{cov}(m_1, m_2) \equiv \overline{\varepsilon_1 \varepsilon_2} &= \frac{[w s_1 s_2]}{\{[w(s_3^2 - s_1^2)]\}\{[w(s_3^2 - s_2^2)]\}} \\
\mathrm{cov}(m_i, d_c) \equiv \overline{\varepsilon_i \varepsilon_3} &= \frac{[w s_i]}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i = 1, 2.
\end{aligned}$$

For the earlier, more general expressions for the components of $\overline{\varepsilon \varepsilon^T}$ it is still necessary to find $\overline{\eta_{ki} \eta_{lj}}$ and $\overline{\eta_{ki} \xi_{l3}}$ in terms of $\overline{\xi_{ki} \xi_{lj}}$, with $\eta_{ki} \equiv \delta s_{ki} = \delta(r_{ki} - \langle r_{ki} \rangle) = \xi_{ki} - \langle \xi_i \rangle = \sum_l w_l (\xi_{ki} - \xi_{li})/[w]$.

$$\begin{aligned}
\overline{\eta_{ki} \eta_{pj}} &= \sum_{l,q} w_l w_q \overline{(\xi_{ki} - \xi_{li})(\xi_{pj} - \xi_{qj})}/[w]^2 \\
&= \overline{(\xi_{ki} - \langle \xi_i \rangle)(\xi_{pj} - \langle \xi_j \rangle)} \\
\overline{\eta_{ki} \xi_{p3}} &= \sum_l w_l \overline{(\xi_{ki} - \xi_{li}) \xi_{p3}}/[w] = \overline{\xi_{ki} \xi_{p3}} - \overline{\langle \xi_i \rangle \xi_{p3}}.
\end{aligned}$$

In the isotropic, 'no-correlation' case, for example, these reduce to

$$\begin{aligned}
\overline{\eta_{ki} \eta_{pi}} &= -w_k \overline{\xi_{ki}^2}/[w] - w_p \overline{\xi_{pi}^2}/[w] \\
&\quad + [w^2 \overline{\xi_i^2}]/[w]^2, \quad k \neq p; \; i = 1, 2 \\
\overline{\eta_{ki}^2} &= (1 - 2w_k/[w]) \overline{\xi_{ki}^2} + [w^2 \overline{\xi_i^2}]/[w]^2, \quad i = 1, 2 \\
\overline{\eta_{k3} \eta_{p3}} &= -w_p \overline{\xi_{p3}^2}/[w],
\end{aligned}$$

and

$$\overline{\eta_{k3}^2} = \overline{\xi_{k3}^2} - w_k \overline{\xi_{k3}^2}/[w] = \overline{\xi_{k3}^2}(1 - w_k/[w]).$$

Here the difference between the correct covariance values and the values obtained on ignoring the variation in $\langle r \rangle$ may be important if the number of defining atoms is small, say, 5 or 4 or, in the extreme, 3.

3.2.2.2. The standard uncertainty of the distance from an atom to the plane

There are two cases, as has been pointed out, *e.g.*, by Ito (1982).

(1) The atom (atom K) was *not* included in the specification of the plane.

$$\begin{aligned}
d_K &= \mathsf{m}^T(\mathsf{r}_K - \langle \mathsf{r} \rangle) = r_{k3} - \langle r_3 \rangle \\
\delta d_K &= \xi_{K3} + s_{K1} \varepsilon_1 + s_{K2} \varepsilon_2 - \varepsilon_3 \\
\sigma_{d_K}^2 &= \overline{\xi_{K3}^2} + s_{K1}^2 \overline{\varepsilon_1^2} + s_{K2}^2 \overline{\varepsilon_2^2} + \overline{\varepsilon_3^2} \\
&\quad + 2 s_{K1} s_{K2} \overline{\varepsilon_1 \varepsilon_2} - 2 s_{K1} \overline{\varepsilon_1 \varepsilon_3} - 2 s_{K2} \overline{\varepsilon_2 \varepsilon_3} \\
&\quad + 2 s_{K1} \overline{\xi_{K3} \varepsilon_1} + 2 s_{K2} \overline{\xi_{K3} \varepsilon_2} - 2 \overline{\xi_{K3} \varepsilon_3}.
\end{aligned}$$

In the isotropic, 'no-correlation' case the last three terms, *i.e.* the terms in $\overline{\varepsilon_i \xi_{K3}}$, are all negligible or zero.

In either case the value for $\overline{\xi_{K3}^2}$ and the appropriate $\overline{\varepsilon_i \varepsilon_j}$ values from the least-squares-plane calculation need to be inserted.

(2) Atom K *was* included in the specification of the plane. The expression for $\sigma_{d_K}^2$ remains the same, but the averages in it may be importantly different.

For example, consider a plane defined by only three atoms, one of overwhelmingly great w at (0, 0, 0), one at (1, 0, 0) and one at (0, 1, 0). The centroid is at (0, 0, 0) and we take $K = 2$, *i.e.* σ_{d_2} is the item of interest. Of course, it is obvious without calculation that the standard uncertainties vanish for the distances of the three atoms from the plane they alone define; the purpose here is only to show, in one case for one of the atoms, that the calculation gives the same result, partly, it will be seen, because the change in orientation of the plane is taken into account. If the only variation in the atom positions is described by $\overline{\xi_{23}^2} = \sigma^2$, one has $s_{21} = 1$, $\varepsilon_3 = \varepsilon_2 = 0$, $\varepsilon_1 = -\xi_{23}$, and $\overline{\xi_{K3} \varepsilon_1} = \sigma^2$. The non-vanishing terms in the desired variance are then

$$\begin{aligned}
\sigma_{d_2}^2 &= \overline{\xi_{23}^2} + s_{21}^2 \overline{\varepsilon_1^2} + 2 s_{21} \overline{\xi_{23} \varepsilon_1} \\
&= (1 + 1 - 2)\sigma^2 = 0.
\end{aligned}$$

If, however, the problem concerns the same plane and a fourth atom at position $(1, 0, r_{43})$, *not* included in the specification of the plane and uncertain only in respect to r_{43} (which is arbitrary) with $\overline{\xi_{43}^2} = \sigma^2$ (the same mean-square variation in direction 3 as for atom 2) and $\overline{\xi_{43} \xi_{23}} = 0$, the calculation for $\sigma_{d_4}^2$ runs the same as before, except for the third term:

$$\sigma_{d_4}^2 = (1 + 1 - 0)\sigma^2 = 2\sigma^2.$$

Extreme examples of this kind show clearly enough that variation in the direction of the plane normal or in the normal component of the centroid position will sometimes be important, the remarks to the contrary by Shmueli (1981) and, for the centroid, the omission by WMC notwithstanding. If only a few atoms are used to define the plane (*e.g.*, three or, as is often the case, a very few more), both the covariance with the centroid position and uncertainty in the direction of the normal are likely to be important. The uncertainty in the normal may still be important, even if a goodly number of atoms are used to define the plane, whenever the test atom lies near or beyond the edge of the lateral domain defined by the other atoms.

3.2.3. The proper least-squares plane, with Gaussian weights

If it is desired to weight the points to be fitted by a plane in the sense of Gaussian least squares, then two items different from what we have seen in the crystallographic literature have to be brought into view: (1) the weights may be anisotropic and include interatomic correlations, because the error matrix of the atom coordinates may in general be anisotropic and include interatomic correlations; and (2) it has to be considered that the atoms truly lie on a plane and that their observed positions are to be adjusted to lie precisely on that plane, whatever its precise position may turn out to be and no matter what the direction, in response to the anisotropic weighting, of their approach to the plane.

An important consequence of (1), the nondiagonal character of the weight matrix, even with Cartesian coordinates, is that the problem is no longer an ordinary eigenvalue problem as treated by SWMB (1959),[1] not even if there is no interatomic correlation and the anisotropy is the same for each atom. On this last case the contrary remark in SWMB at the beginning of the footnote, p. 601, is incorrect, and the treatments in terms of the eigenvector–eigenvalue problem by Hamilton (1961, 1964, pp. 174–177) and Ito (1981*a*)[2] evade us. At best the problem is still not a genuine eigenvalue problem if the anisotropies of the atoms are not all alike.

Hypothesis (2), of perfect planarity, may be hard to swallow. It has to be tested, and for any set of atoms the conclusion may be that they probably do not lie on a plane. But if the hypothesis is provisionally adopted (and it has to be decided beforehand which of the following alternatives is to be followed), the adjusted positions are obtained by moving the atoms onto the plane

(*a*) along paths normal to the plane, or

(*b*) along the *proper* paths of 'least resistance' – that is, paths with, in general, both normal and lateral components differently directed for each atom so as to minimize the appropriately weighted quadratic form of differences between the observed and adjusted coordinates. The lateral motions (and the anisotropic weights that induce them) may change the relative weights of different atoms in accordance with the orientation of the plane; change the perpendicular distance of origin-to-plane; and change the orientation of the plane in ways that may at first give surprise.

Our first example of this has already been given.[3] A second, which actually inspired our first, is to be found in Hamilton (1964; example 5-8-1, p. 177), who discusses a rectangle of four points with identical error ellipsoids elongated in the long direction of the rectangle. The unweighted best line bisects the pattern along this direction, but the weighted best line is parallel to the short direction, if the elongation of the ellipsoids is sufficient. A third example (it is severely specialized so that a precise result can be attained without calculation) has three atoms ABC arranged like a $C_{2v}-mm$ molecule with bond angle 90°. The central atom, B, has overwhelming isotropic weight; A and C have parallel extremely elongated error ellipsoids, aligned parallel to the $A-B$ bond. The unweighted best line passes through B parallel to $A\cdots C$; the weighted best line passes through B and through C. Our last example is of a plane defined by a number of atoms of which one lies a considerable distance above the plane and at a distance from the normal through the centroid, but with the downward semi-axis of its extremely elongated prolate error ellipsoid intersecting that normal before it intersects the plane. If this atom is moved at right angles to the plane and further away from it, the centroid normal tips toward the atom, whereas it would tip away if the atom's weight function were isotropic or if the calculation were the usual one and in effect constrained the adjusted position of the atom to move at right angles to the plane.

The lead notion here – that the observed points are to be adjusted individually to fit a curve (surface) of required type exactly, rather than that the curve should simply be constructed to fit the observed points as well as possible in the sense of minimizing the weighted sum of squares of the distances along some preordained direction (perhaps normal to the plane, but perhaps as in ordinary curve fitting parallel to the y axis) – we first learned from the book by Deming (1943), *Statistical Adjustment of Data*, but it is to be found in Whittaker & Robinson (1929), Arley & Buch (1950), Hamilton (1964, our most used reference), and doubtless widely throughout the least-squares literature. It has recently been strongly emphasized by Lybanon (1984), who gives a number of modern references. It is the only prescription that properly satisfies the least-squares conditions, whereas (*a*) and other analogous prescriptions are only arbitrary regressions, in (*a*) a normal regression onto the plane.[4]

We have explored the problem of least-squares *adjustment* of observed positions subject to anisotropic weights with the help of three Fortran programs, one for the straight line and two for the plane. In the first plane program an approximate plane is derived, coordinates are rotated as in WMC (1973), and the parameters of the plane are adjusted and the atoms moved onto it, either normally or in full accord with the least-squares condition, but in either case subject to independent anisotropic weight matrices. The other plane program, described in Appendix A3.2.1, proceeds somewhat more directly, with the help of the method of Lagrange multipliers. However, neither program has been brought to a satisfactory state for the calculation of the variances and covariances of the derived quantities.

3.2.3.1. Formulation and solution of the general Gaussian plane

We conclude with an outline for a complete feasible solution, including interatomic weight-matrix elements. Consider atoms at observed vector positions $\mathbf{r}_k$, $k = 1, \ldots, n$, designated in the following equations by $\mathbf{R}$, an n-by-3 array, with $\boldsymbol{R}_{ki} = \boldsymbol{r}_{ki}$; the corresponding adjusted positions denoted by the array $\mathbf{R}_a$; n constraints (each adjusted position $\mathbf{r}_a$ – 'a' for 'adjusted' – must be

[1] A simple two-dimensional problem illustrates the point. A regular polygon of n atoms is to define a 'best' line (always a *central* line). If the error matrix (the same for each atom) is isotropic, the weighted sum of squares of deviations from the line is independent of its orientation for $n > 2$, *i.e.* the problem is a degenerate eigenvalue problem, with two equal eigenvalues. However, if the error ellipsoids are not isotropic and are all oriented radially or all tangentially (these are merely the two orientations tried), the sum has $n/2$ equal minima for even n and 2 equal minima for odd n, in the one-π range of possible orientations of the line. Possibly similar peculiarities might be imagined if the anisotropic weights were more complicated (*e.g.*, 'star' shaped) than can be described by a nonsingular matrix, or by any matrix. Such are of course excluded here.

[2] Ito observes that his method fails when there are only three points to define the plane, his least-squares normal equations becoming singular. But the situation is worse: his equations are singular for any number of points, if the points fit a plane exactly.

[3] See first footnote[1].

[4] Ito's second method (Ito, 1981*b*), of 'substitution', is also a regression, essentially like the regression along z at fixed x and y used long ago by Clews & Cochran (1949, p. 52) and like the regressions of y on fixed x that – despite the fact that both x and y are afflicted with random errors – are commonly taught or practised in schools, universities and laboratories nearly 200 years after Gauss, to the extent that Deming, Lybanon and other followers of Gauss have so far had rather little influence. Kalantar's (1987) short note is a welcome but still rare exception.

on the plane), and $3n+3$ adjustable parameters ($3n$ $\mathbf{R}_a$ components and the 3 components of the vector $\mathbf{b}$ of reciprocal intercepts of the plane), so that the problem has $n-3$ degrees of freedom. The $3n$-by-$3n$ weight matrix $\boldsymbol{P}$ may be anisotropic for the separate atoms, and may include interatomic elements; it is symmetric in the sense $\boldsymbol{P}_{kilj} = \boldsymbol{P}_{ljki}$, but $\boldsymbol{P}_{kilj}$ will not in general be equal to $\boldsymbol{P}_{kjli}$. The array $\boldsymbol{\Lambda}$ denotes n Lagrange multipliers, one for each atom and unrelated to the λ's of Section 3.2.2; m and d still represent the direction cosines of the plane normal and the perpendicular origin-to-plane distance.

For a linear least-squares problem we know (see, *e.g.*, Hamilton, 1964, p. 143) that the proper weight matrix is the reciprocal of the atomic error matrix ($\boldsymbol{P} = \boldsymbol{M}^{-1}$);[5] note that '$\boldsymbol{M}$' is unrelated to the '$\mathbf{M}$' of Section 3.2.2. The least-squares sum is now

$$S = (\mathbf{R} - \mathbf{R}_a)\boldsymbol{P}(\mathbf{R} - \mathbf{R}_a),$$

and the augmented sum for applying the method of Lagrange multipliers is

$$\Phi = S/2 - \boldsymbol{\Lambda}\mathsf{b}\mathbf{R}_a.$$

Here the summation brackets ([...]) have been replaced by matrix products, but the simplified notation is not matrix notation. It has to be regarded as only a useful mnemonic in which all the indices and their clutter have been suppressed in view of their inconveniently large number for some of the arrays. If the dimensions of each are kept in mind, it is easy to recall the indices, and if the indices are resupplied at any point, it is not difficult to discover what is really meant by any of the expressions or whether evaluations have to be executed in a particular order.

The conditions to be satisfied are

$$\begin{aligned} 0 &= \frac{-\partial\Phi}{\partial\mathbf{R}_a} = \boldsymbol{P}(\mathbf{R} - \mathbf{R}_a) + \boldsymbol{\Lambda}\mathsf{b}, \\ 0 &= \frac{-\partial\Phi}{\partial\mathsf{b}} = \boldsymbol{\Lambda}\mathbf{R}_a, \\ \mathbf{1} &= \mathsf{b}\mathbf{R}_a. \end{aligned}$$

That the partial derivatives of $S/2$ should be represented by $\boldsymbol{P}(\mathbf{R} - \mathbf{R}_a)$ depends upon the above-mentioned symmetry of $\boldsymbol{P}$. Note that each of the n unit elements of $\mathbf{1}$ expresses the condition that its $\mathbf{r}_a$ should indeed lie on the plane, and that $\boldsymbol{\Lambda}\mathsf{b}$ is just the same as $\mathsf{b}\boldsymbol{\Lambda}$. The perpendicular origin-to-plane distance and the direction cosines of the plane normal are $d^2 = 1/\mathsf{b}^T\mathsf{b}$ and $\mathsf{m} = \mathsf{b}/\sqrt{\mathsf{b}^T\mathsf{b}}$.

On multiplication by $\boldsymbol{M}$ the first condition solves for $\mathbf{R}_a$, and that expression combined separately with the second condition and with the third gives highly nonlinear equations (there is further mention of this nonlinearity in Appendix A3.2.1) that have to be solved for b and $\boldsymbol{\Lambda}$:

$$\begin{aligned} \mathbf{R}_a &= \mathbf{R} + \boldsymbol{M}\boldsymbol{\Lambda}\mathsf{b} = \mathbf{R} + \boldsymbol{M}\mathsf{b}\boldsymbol{\Lambda} \\ 0 &= \mathbf{F} \equiv (\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda})\mathsf{b} + \boldsymbol{\Lambda}\mathbf{R} \\ 0 &= \mathbf{G} \equiv (\mathsf{b}\boldsymbol{M}\mathsf{b})\boldsymbol{\Lambda} - (\mathbf{1} - \mathsf{b}\mathbf{R}). \end{aligned}$$

The best way of handling these equations, at least if it is desired to find their solutions both for $\mathbf{R}_a$, $\boldsymbol{\Lambda}$, and b and for the error matrix of these derived quantities in terms of the error matrix for $\mathbf{R}$, seems to be to find an approximate solution by one means or another, to make sure that it is the desired solution, if (as may be) there is ambiguity, and to shift the origin to a point far enough from the plane and close enough to the centroid normal to avoid the difficulties discussed by SWMB. Then linearize the first and second equations by differentiation and solve the results first iteratively to fit the current residuals $\mathbf{F}_0$ and $\mathbf{G}_0$ and then for nominally infinitesimal increments $\delta\mathsf{b}$ and $\delta\boldsymbol{\Lambda}$. In effect, one deals with equations $\mathbf{Q}\mathbf{X} = \mathbf{Y}$, where $\mathbf{Q}$ is the $(n+3)\times(n+3)$ matrix of coefficients of the following set of equations, $\mathbf{X}$ is the $(n+3)$-dimensional vector of increments $\delta\mathsf{b}$ and $\delta\boldsymbol{\Lambda}$, and $\mathbf{Y}$ is the vector either of the first terms or of the second terms on the right-hand side of the equations.

$$\begin{aligned} (\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda})\delta\mathsf{b} + (\boldsymbol{M}\boldsymbol{\Lambda}\mathsf{b} + \boldsymbol{\Lambda}\boldsymbol{M}\mathsf{b} + \mathbf{R})\delta\boldsymbol{\Lambda} &= -\mathbf{F}_0 - \boldsymbol{\Lambda}\delta\mathbf{R} \\ (\boldsymbol{M}\mathsf{b}\boldsymbol{\Lambda} + \mathsf{b}\boldsymbol{M}\boldsymbol{\Lambda} + \mathbf{R})\delta\mathsf{b} + (\mathsf{b}\boldsymbol{M}\mathsf{b})\delta\boldsymbol{\Lambda} &= -\mathbf{G}_0 - \mathsf{b}\delta\mathbf{R}. \end{aligned}$$

When $\mathbf{X}$ becomes the vector ε of errors in b and $\boldsymbol{\Lambda}$ as induced by errors $\eta \equiv \delta\mathbf{R}$ in the measured atom positions, these equations become, in proper matrix notation, $\mathbf{Q}\varepsilon = \mathbf{Z}\eta$, with solution $\varepsilon = \mathbf{Q}^{-1}\mathbf{Z}\eta$, where $\mathbf{Z}$ is the $(n+3)$-dimensional vector of components, first of b then of $\boldsymbol{\Lambda}$. The covariance matrix $\overline{\varepsilon\varepsilon^T}$, from which all the covariances of b, $\mathbf{R}_a$, and $\mathbf{R}$ (including for the latter any atoms that were not included for the plane) can be derived, is then given by

$$\overline{\varepsilon\varepsilon^T} = \mathbf{Q}^{-1}\mathbf{Z}\overline{\eta\eta^T}\mathbf{Z}^T(\mathbf{Q}^{-1})^T.$$

This is not as simple as the familiar expression for propagation of errors for a least-squares problem solved without the use of Lagrange multipliers, *i.e.* $\overline{\varepsilon\varepsilon^T} = \mathbf{B}^{-1}$, where $\mathbf{B}$ is the matrix of the usual normal equations, both because $\mathbf{B} = \mathbf{B}^T$ is no longer valid and because the middle factor $\mathbf{Z}\overline{\eta\eta^T}\mathbf{Z}^T$ is no longer equal to $\mathbf{B}^{-1}$.

It is easy to verify that $\boldsymbol{\Lambda}$ consists of a set of coefficients for combining the n row vectors of $\boldsymbol{M}\mathsf{b}$, in the expression for $\mathbf{R}_a$, into corrections to $\mathbf{R}$ such that each adjusted position lies exactly on the plane:

$$\begin{aligned} \mathsf{b}\mathbf{R}_a &= \mathsf{b}\mathbf{R} + \mathsf{b}\boldsymbol{M}\mathsf{b}(\mathsf{b}\boldsymbol{M}\mathsf{b})^{-1}(\boldsymbol{1} - \mathsf{b}\mathbf{R}) \\ &= \mathsf{b}\mathbf{R} + \boldsymbol{1} - \mathsf{b}\mathbf{R} = \boldsymbol{1}. \end{aligned}$$

At the same time one can see how the lateral shifts occur in response to the anisotropy of $\boldsymbol{M}$, especially if, now, only the anisotropic case without interatomic correlations is considered. For atom k write b in terms of its components along the principal axes of $\mathbf{M}_k$, associated with the eigenvalues α, β and γ; the shifts are then proportional to $\mathbf{M}_{\alpha\alpha}b_\alpha^2$, $\mathbf{M}_{\beta\beta}b_\beta^2$ and $\mathbf{M}_{\gamma\gamma}b_\gamma^2$, each along its principal axis, and the sums of the contributions of these shift components to the total displacement along the plane normal or along either of two orthogonal directions in the plane can readily be visualized. In effect $\mathbf{M}_k$ is the compliance matrix for these shifts of atom k. Similarly, it can be seen that in the isotropic case with interatomic correlations a pair of equally weighted atoms

[5] Is this statement firm for a nonlinear problem? We use it, assuming that at convergence the problem has become effectively linear. But in fact this will depend on how great the nonlinearity is, in comparison with the random errors (variances) that eventually have to be considered. Another caveat may be in order in regard to our limited knowledge of Gauss's second derivation of the method of least squares, the one he preferred [see Whittaker & Robinson (1929)] and which establishes for a linear system that the best *linear combination* of a set of observations, afflicted by random errors, for estimating any arbitrary derived quantity – best in the sense of being unbiased and having minimal mean-square error – is given by the method of least squares with the weight matrix set equal to the inverse error matrix of the observations. Hamilton, and Whittaker & Robinson, prove this only for the case that the derived parameters are not constrained, whereas here they are. We believe, however, that the best choice of weights is a question concerning only the observations, and that it cannot be affected by the method used for minimizing $\boldsymbol{S}$ subject to any constraints, whether by eliminating some of the parameters by invoking the constraints directly or by the use of Lagrange multipliers.

located, for example, at some distance apart and at about the same distance from the plane, will have different shifts (and different influences on d and m) depending on whether the covariance between the errors in the perpendicular components of their observed positions relative to the plane is small, or, if large, whether it is positive or is negative. If the covariance is large and positive, the adjusted positions will both be pulled toward the plane, strongly resisting, however, the apparent necessity that both reach the plane by moving by different amounts; in consequence, there will be a strong tendency for the plane to tilt toward the more distant atom, and possibly even away from the nearer one. But if the covariance is large and negative, the situation is reversed: the more distant atom can readily move more than the nearer one, while it is very difficult to move them together; the upshot is then that the plane will move to meet the original midpoint of the two atoms while they tilt about that midpoint to accommodate the plane.

It is attractive to solve our problem with the 'normal' formulation of the plane, $\mathsf{mr} = d$, and so directly avoid the problems that arise for $d \approx 0$. The solution in terms of the reciprocal intercepts b has been given first, however, because reducing by two (d and a Lagrange multiplier) the number of parameters to be solved for may more than make up for the nuisance of having to move the origin. The solution in terms of d follows.

The augmented variation function is now

$$\Phi = (\mathbf{R} - \mathbf{R}_a)\boldsymbol{P}(\mathbf{R} - \mathbf{R}_a)/2 - \boldsymbol{\Lambda}(\mathsf{m}\mathbf{R}_a - d\boldsymbol{1}) + \mu\mathsf{m}\mathsf{m}/2,$$

the term in the new Lagrange multiplier, μ, and the term in $d\boldsymbol{1}$ having been added to the previous expression. The $\boldsymbol{1}$, an n-vector of 1's, is needed to express the presence of n terms in the $\boldsymbol{\Lambda}$ sum. There are then five equations to be satisfied – actually $n + 1 + 3n + 3 + 1 = 4n + 5$ ordinary equations – in the $3n$ $\mathbf{R}_a$ components, the n $\boldsymbol{\Lambda}$'s, the 3 m components, μ, and $d - 4n + 5$ unknowns in all, as required. These equations are as follows:

$$\begin{aligned}
\mathsf{m}\mathbf{R}_a &= d\boldsymbol{1} \\
\mathsf{m}\mathsf{m} &= 1 \\
0 &= -\left(\frac{\partial\Phi}{\partial\mathbf{R}_a}\right) = \boldsymbol{P}(\mathbf{R} - \mathbf{R}_a) + \boldsymbol{\Lambda}\mathsf{m} \\
0 &= \frac{\partial\Phi}{\partial\mathsf{m}} = -\boldsymbol{\Lambda}\mathbf{R}_a + \mu\mathsf{m} \\
0 &= \frac{\partial\Phi}{\partial d} = \boldsymbol{\Lambda}\boldsymbol{1}.
\end{aligned}$$

As before, multiply the third equation by $\boldsymbol{M}$ and solve for $\mathbf{R}_a$. Then substitute the result into the first and fourth equations to obtain

$$\begin{aligned}
\mathsf{m}\mathbf{R} + \mathsf{m}\boldsymbol{M}\mathsf{m}\boldsymbol{\Lambda} &= d\boldsymbol{1}, \\
\mathsf{m}\mathsf{m} &= 1, \\
\boldsymbol{\Lambda}\mathbf{R} + \boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda}\mathsf{m} &= \mu\mathsf{m}, \\
\boldsymbol{\Lambda}\boldsymbol{1} &= 0
\end{aligned}$$

as the $n + 5$ mostly nonlinear equations to be solved for m, $\boldsymbol{\Lambda}$, d and μ by linearizing (differentiation), solving for increments, and iterating, in the pattern described more fully above. An approximate solution for m and d has first to be obtained somehow, perhaps by the method of SWMB (with isotropic uncorrelated weights), checked for suitability, and extended to a full complement of first approximations by

$$\begin{aligned}
\boldsymbol{\Lambda} &= (\mathsf{m}\boldsymbol{M}\mathsf{m})^{-1}(d\boldsymbol{1} - \mathsf{m}\mathbf{R}) \\
\mu &= \mathsf{m}\boldsymbol{\Lambda}\mathbf{R} + \mathsf{m}\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda}\mathsf{m},
\end{aligned}$$

which readily follow from the previous equations. As in the 'intercepts' solution the linearized expression for the increments in μ, $\boldsymbol{\Lambda}$, d and m can be used together with the equation for $\mathbf{R}_a$ to obtain all the covariances needed in the treatment described in Section 3.2.2.

3.2.3.2. Concluding remarks

Proper tests of statistical significance of this or that aspect of a least-squares plane can be made if the plane has been based on a proper weight matrix as discussed in Section 3.2.3; if it can be agreed that the random errors of observation are normally distributed; and if an agreeable test (*null*) hypothesis can be formulated. For example, one may ask for the probability that a degree of fit of the observed positions to the least-squares plane at least as poor as the fit that was found might occur if the atoms in truth lie precisely on a plane. The χ^2 test answers this question: a table of probabilities displayed as a function of χ^2 and ν provides the answer. Here χ^2 is just our minimized

$$S = \mathsf{b}\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{P}\boldsymbol{M}\boldsymbol{\Lambda}\mathsf{b} = \mathsf{b}\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda}\mathsf{b},$$

and

$$\begin{aligned}
\nu &= n_{\text{observations}} - n_{\text{adjusted parameters}} - n_{\text{constraints}} \\
&= 3n - (n + 3) - n = n - 3,
\end{aligned}$$

is the number of degrees of freedom for the problem of the plane (erroneously cited in at least one widely used crystallographic system of programs as $3n - 3$). There will not usually be any reason to believe that the atoms are exactly coplanar in any case; nevertheless, this test may well give a satisfying indication of whether or not the atoms are, in the investigator's judgment, essentially coplanar. It must be emphasized that χ^2 as calculated in Section 3.2.3 will include proper allowance for uncertainties in the d and orientation of the plane with greater reliability than the estimates of Section 3.2.2, which are based on nominally arbitrary weights. Both, however, will allow for the large variations in d and *tilt* that can arise in either case if n is small. Some of the earlier, less complete discussions of this problem have been mentioned in Section 3.2.2.

Among the problems not considered here are ones of fitting more than one plane to a set of observed positions, *e.g.* of two planes fitted to three sets of atoms associated, respectively, with the first plane, the second plane, and both planes, and of the angle between the two planes. For the atoms common to both planes there will be a fundamental point of difference between existing programs (in which, in effect, the positions of the atoms in common are inconsistently adjusted to one position on the first plane and, in general, a different position on the second) and what we would advocate as the proper procedure of requiring the adjusted positions of such atoms to lie on the line of intersection of the two planes. As to the dihedral angle there is a difficulty, noted by WMC (1973, p. 2705), that the usual formulation of $\sigma^2(\theta_0)$ in terms of the cosine of the dihedral angle reduces to 0/0 at $\theta_0 = 0$. However, this variance is obviously well defined if the plane normals and their covariances are well defined. The essential difficulty lies with the ambiguity in the direction of the line of intersection of the planes in the limit of zero dihedral angle. For the torsion angle about a line defined by two atoms, there should be no such difficulty. It seems likely that for the two-plane problem proposed above, the issue that decides whether

the dihedral angle will behave like the standard dihedral angle or, instead, like the torsion angle, will be found to be whether or not two or more atoms are common to both planes.

All that we have tried to bring out about the covariances of derived quantities involving the plane requires that the covariances of the experimental atom positions (reduced in our formulations to Cartesian coordinates) be included. However, such covariances of derived quantities are often not available in practice, and are usually left unused even if they are. The need to use the covariances, not just the variances, has been obvious from the beginning. It has been emphasized in another context by Schomaker & Marsh (1983) and much more strongly and generally by Waser (1973), whose pleading seems to have been generally ignored, by now, for about thirty years.

APPENDIX A3.2.1

Consider n atoms at observed vector positions r (expressed in Cartesians), n constraints (each adjusted position r_a – 'a' for 'adjusted' – must be on the plane) and $3n+3$ adjustable parameters ($3n$ r_a components and the 3 components of the vector a of reciprocal intercepts of the plane), so that the problem has $n-3$ degrees of freedom. The weight matrices $\mathbf{P}$ may be differently anisotropic for each atom, but there are no interatomic correlations. As before, square brackets, '[...]', represent the Gaussian sum over all atoms, usually suppressing the atom indices. We also write λ, *not* the λ of Section 3.2.2, for the Lagrange multipliers (one for each atom); m for the direction cosines of the plane normal; and d for the perpendicular origin-to-plane distance.

As before, $\mathbf{P}_k$ is the reciprocal of the atomic error matrix: $\mathbf{P}_k = \mathbf{M}_k^{-1}$ (correspondingly, $\boldsymbol{P} \equiv \boldsymbol{M}^{-1}$) but '$\mathbf{M}$' is no longer the '$\mathbf{M}$' of Section 3.2.2. The appropriate least-squares sum is

$$S = [(\mathsf{r} - \mathsf{r}_a)^T \mathbf{P} (\mathsf{r} - \mathsf{r}_a)]$$

and the augmented sum for applying the method of Lagrange multipliers is

$$\Phi = S/2 - [\lambda \mathsf{a}^T \mathsf{r}_a].$$

Φ is to be minimized with respect to variations of the adjusted atom positions r_{ka} and plane reciprocal intercepts b_i, leading to the equations

$$0 = \frac{-\partial \phi}{\partial \mathsf{r}_a^T} = \mathbf{P}(\mathsf{r} - \mathsf{r}_a) + \lambda \mathsf{b} \text{ and}$$
$$0 = \frac{-\partial \Phi}{\partial \mathsf{b}^T} = [\lambda \mathsf{r}_a],$$

subject to the plane conditions $\mathsf{b}^T \mathsf{r}_a = 1$, each atom, with $d^2 = 1/(\mathsf{b}^T \mathsf{b})$, $\mathsf{m} = \mathsf{b}/\sqrt{\mathsf{b}^T \mathsf{b}}$. These equations are nonlinear.

A convenient solution runs as follows: first multiply the first equation by $\mathbf{M}$ and solve for the adjusted atom positions in terms of the Lagrange multipliers λ and the reciprocal intercepts b of the plane; then multiply *that* result by b^T applying the plane conditions, and solve for the λ's

$$\mathsf{r}_a = \mathsf{r} + \lambda \mathbf{M} \mathsf{b}, \quad \mathbf{M} \equiv \mathbf{P}^{-1}$$
$$1 = \mathsf{b}^T \mathsf{r} + \lambda \mathsf{b}^T \mathbf{M} \mathsf{b},$$
$$\lambda = (1 - \mathsf{b}^T \mathsf{r})/(\mathsf{b}^T \mathbf{M} \mathsf{b}).$$

Next insert these values for the λ's and r_a's into the second equation:

$$0 = \frac{\partial \Phi}{\partial \mathsf{b}^T} = [\lambda \mathsf{r}_a] = \left[\frac{1 - \mathsf{b}^T \mathsf{r}}{\mathsf{b}^T \mathbf{M} \mathsf{b}} \left(\mathsf{r} + \frac{1 - \mathsf{b}^T \mathsf{r}}{\mathsf{b}^T \mathbf{M} \mathsf{b}} \mathbf{M} \mathsf{b} \right) \right].$$

This last equation, $F(\mathsf{b}) = 0$, is to be solved for b. It is highly nonlinear: $F(\mathsf{b}) = O(\mathsf{b}^3)/O(\mathsf{b}^4)$. One can proceed to a first approximation by writing $0 = [(1 - \mathsf{b}^T \mathsf{r})\mathsf{r}]$; *i.e.*, $[\mathbf{rr}] \cdot \mathsf{b} = [\mathbf{r}]$, in dyadic notation. [$\mathbf{M} = \mathbf{I}$, all atoms; $1 - \mathsf{b}^T \mathsf{r} = 0$ in the multiplier of $\mathbf{M}\mathsf{b}/(\mathsf{b}^T \mathbf{M} \mathsf{b})$.] A linear equation in b, this approximation usually works well.[6] We have also used the iterative Frazer, Duncan & Collar eigenvalue solution as described by SWMB (1959), which works even when the plane passes exactly through the origin. To continue the solution of the nonlinear equations, by linearizing and iterating, write $F(\mathsf{b}) = 0$ in the form

$$0 = [\lambda \mathsf{r} + \lambda^2 \mathbf{M} \mathsf{b}]$$
$$\approx \left[\mathsf{r} \frac{\partial \lambda}{\partial \mathsf{b}} + 2 \mathbf{M} \mathsf{b} \lambda \frac{\partial \lambda}{\partial \mathsf{b}} + \lambda^2 \mathbf{M} \right] \delta \mathsf{b} + [\lambda(\mathsf{r} + \lambda \mathbf{M} \mathsf{b})]_0,$$

solve for $\delta\mathsf{b}$, reset b to $\mathsf{b} + \delta\mathsf{b}$, *etc.,* until the desired degree of convergence of $|\delta\mathsf{b}|/|\mathsf{b}|$ toward zero has been attained. By $\partial\lambda/\partial\mathsf{b}$ is meant the partial derivative of the above expression for λ with respect to b, as detailed in the next paragraph.

In the Fortran program *DDLELSP* (double precision Deming Lagrange, with error estimates, least-squares plane, written to explore this solution) the preceding equation is recast as

$$\mathbf{B} \delta \mathsf{b} \equiv \left[\lambda^2 \mathbf{M} + (\mathsf{r} + 2\lambda \mathbf{M} \mathsf{b}) \frac{\partial \lambda}{\partial \mathsf{b}} \right] \delta \mathsf{b}$$
$$\equiv \mathbf{Y} = -[\lambda(\mathsf{r} + \lambda \mathbf{M} \mathsf{b})]_0,$$

with

$$\frac{\partial \lambda}{\partial \mathsf{b}} = \mathsf{r}^T / \mathsf{b}^T \mathbf{M} \mathsf{b} - [2(1 - \mathsf{b}^T \mathsf{r})/(\mathsf{b}^T \mathbf{M} \mathsf{b})^2]$$
$$= -(\mathsf{r}^T + 2\lambda \mathsf{b}^T \mathbf{M})/\mathsf{b}^T \mathbf{M} \mathsf{b}.$$

The usual goodness of fit, GOF2 in *DDLELSP*, evaluates to

$$\text{GOF2} = \left(\frac{S_{\min}}{n-3} \right)^{1/2} = \left(\frac{1}{n-3} [\lambda^2 \mathsf{b}^T \mathbf{M} \mathbf{P} \mathbf{M} \mathsf{b}] \right)^{1/2}$$
$$= \left(\frac{1}{n-3} [\lambda^2 \mathsf{b}^T \mathbf{M} \mathsf{b}] \right)^{1/2}$$
$$= \left(\frac{1}{n-3} \left[\frac{(1 - \mathsf{b}^T \mathsf{r})^2}{\mathsf{b}^T \mathbf{M} \mathsf{b}} \right] \right)^{1/2}.$$

[6] We do not fully understand the curious situation of this equation. It arises immediately if the isotropic problem is formulated as one of minimizing $[(1 - \mathsf{b}^T \mathsf{r})^2]$ by varying b, and it fails then [SWMB (1959) referred to it as 'an incorrect method'], as it obviously must – observe the denominator – if the plane passes too close to the origin. However, it fails in other circumstances also. The main point about it is perhaps that it is linear in b and is obtained as the supposedly exact and *unique* solution of the isotropic problem, whereas the problem has no unique solution but three solutions instead (SWMB, 1959). From the point of view of Gaussian least squares, the essential fault in minimizing $S_{\text{lin}} = [(1 - \mathsf{b}^T \mathsf{r})^2]$ may be that the apparently simple weighting function in it, *i.e.* the identity, is actually complicated and unreasonable. In terms of distance deviations from the plane, we have $S_{\text{lin}} = [w(d - \mathsf{m}^T \mathsf{r})^2]$, with $w = \mathsf{b}^T \mathsf{b} = d^{-2}$. Prudence requires that the origin be shifted to a point sufficiently far from the plane and close enough to the centroid normal to avoid the difficulties discussed by SWMB. Note that for the one-dimensional problem of fitting a constant to a set of measurements of a single entity the Deming–Lagrange treatment with the condition $1 = cx_a$ and weights w reduces immediately to the standard result $1/c = [wx]/[w]$.

This is only an approximation, because the residuals $1 - \mathsf{b}^T\mathsf{r}$ are not the differences between the observations and appropriate linear functions of the parameters, nor are their variances (the $\mathsf{b}^T\mathbf{M}\mathsf{b}$'s) independent of the parameters (or, in turn, the errors in the observations).

We ask also about the perpendicular distances, e, of atoms to plane and the mean-square deviation $\overline{(\delta e)^2}$ to be expected in e.

$$e = (1 - \mathsf{b}^T\mathsf{r})/\sqrt{\mathsf{b}^T\mathsf{b}} = d(1 - \mathsf{b}^T\mathsf{r})$$
$$\delta e = -d(\mathsf{b}^T\eta + \mathsf{r}^T\varepsilon) + O(\eta^2).$$

Here η and ε are the errors in r and b. Neglecting '$O(\eta^2)$' then leads to

$$\overline{(\delta e)^2} = d^2(\mathsf{b}^T\overline{\eta\eta^T}\mathsf{b} + 2\mathsf{b}^T\overline{\eta\varepsilon^T}\mathsf{r} + \mathsf{r}^T\overline{\varepsilon\varepsilon^T}\mathsf{r}).$$

We have $\overline{\eta\eta^T} = \mathbf{M} = \mathbf{P}^{-1}$, but $\overline{\eta\varepsilon^T}$ and $\overline{\varepsilon\varepsilon^T}$ perhaps still need to be evaluated.

References

Arley, N. & Buch, K. R. (1950). *Introduction to the Theory of Probability and Statistics.* New York: John Wiley and London: Chapman and Hall.

Clews, C. J. B. & Cochran, W. (1949). *The structures of pyrimidines and purines. III. An X-ray investigation of hydrogen bonding in aminopyrimidines. Acta Cryst.* **2**, 46–57.

Deming, W. E. (1943). *Statistical Adjustment of Data.* New York: John Wiley.

Hamilton, W. C. (1961). *On the least-squares plane through a set of points. Acta Cryst.* **14**, 185–189.

Hamilton, W. C. (1964). *Statistics in Physical Science.* New York: Ronald Press.

Ito, T. (1981*a*). *Least-squares refinement of the best-plane parameters. Acta Cryst.* A**37**, 621–624.

Ito, T. (1981*b*). *On the least-squares plane through a group of atoms. Sci. Pap. Inst. Phys. Chem. Res. Saitama*, **75**, 55–58.

Ito, T. (1982). *On the estimated standard deviation of the atom-to-plane distance. Acta Cryst.* A**38**, 869–870.

Kalantar, A. H. (1987). *Slopes of straight lines when neither axis is error-free. J. Chem. Educ.* **64**, 28–29.

Lybanon, M. (1984). *A better least-squares method when both variables have uncertainties. Am. J. Phys.* **52**, 22–26.

Robertson, J. M. (1948). *Bond-length variations in aromatic systems. Acta Cryst.* **1**, 101–109.

Schomaker, V. & Marsh, R. E. (1983). *On evaluating the standard deviation of U_{eq}. Acta Cryst.* A**39**, 819–820.

Schomaker, V., Waser, J., Marsh, R. E. & Bergman, G. (1959). *To fit a plane or a line to a set of points by least squares. Acta Cryst.* **12**, 600–604.

Shmueli, U. (1981). *On the statistics of atomic deviations from the 'best' molecular plane. Acta Cryst.* A**37**, 249–251.

Waser, J. (1973). *Dyadics and variances and covariances of molecular parameters, including those of best planes. Acta Cryst.* A**29**, 621–631.

Waser, J., Marsh, R. E. & Cordes, A. W. (1973). *Variances and covariances for best-plane parameters including dihedral angles. Acta Cryst.* B**29**, 2703–2708.

Whittaker, E. T. & Robinson, G. (1929). *The Calculus of Observations.* London: Blackie.

International Tables for Crystallography (2010). Vol. B, Chapter 3.3, pp. 418–448.

3.3. Molecular modelling and graphics

By R. Diamond and L. M. D. Cranswick

3.3.1. Graphics

By R. Diamond

3.3.1.1. Coordinate systems, notation and standards

3.3.1.1.1. Cartesian and crystallographic coordinates

It is usual, for purposes of molecular modelling and of computer graphics, to adopt a Cartesian coordinate system using mutually perpendicular axes in a right-handed system using the ångström unit or the nanometre as the unit of distance along such axes, and largely to ignore the existence of crystallographic coordinates expressed as fractions of unit-cell edges. Transformations between the two are thus associated, usually, with the input and output stages of any software concerned with modelling and graphics, and it will be assumed after this section that all coordinates are Cartesian using the chosen unit of distance as the unit of coordinates. For a discussion of coordinate transformations and rotations without making this assumption see Chapter 1.1 in which formulations using co- and contravariant forms are presented.

The relationship between these systems may be written

$$\mathbf{X} = \boldsymbol{M}\mathbf{x} \quad \mathbf{x} = \boldsymbol{M}^{-1}\mathbf{X}$$

in which $\mathbf{X}$ and $\mathbf{x}$ are position vectors in direct space, written as column vectors, with $\mathbf{x}$ expressed in crystallographic fractional coordinates (dimensionless) and $\mathbf{X}$ in Cartesian coordinates (dimension of length).

There are two forms of $\boldsymbol{M}$ in common use. The first of these sets the first component of $\mathbf{X}$ parallel to $\mathbf{a}^*$ and the third parallel to $\mathbf{c}$ and is

$$\boldsymbol{M} = \begin{pmatrix} a\varphi/\sin\alpha & 0 & 0 \\ a(\cos\gamma - \cos\alpha\cos\beta)/\sin\alpha & b\sin\alpha & 0 \\ a\cos\beta & b\cos\alpha & c \end{pmatrix}$$

$$\boldsymbol{M}^{-1} = \begin{pmatrix} \sin\alpha/a\varphi & 0 & 0 \\ (\cos\alpha\cos\beta - \cos\gamma)/b\varphi\sin\alpha & 1/b\sin\alpha & 0 \\ (\cos\alpha\cos\gamma - \cos\beta)/c\varphi\sin\alpha & -1/c\tan\alpha & 1/c \end{pmatrix}$$

in which

$$\varphi = \sqrt{1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma} = \sin\alpha\sin\beta\sin\gamma^*.$$

φ is equal to the volume of the unit cell divided by abc, and is unchanged by cyclic permutation of α, β and γ and of α^*, β^* and γ^*. The Cartesian and crystallographic axes have the same chirality if the positive square root is taken.

The second form sets the first component of $\mathbf{X}$ parallel to $\mathbf{a}$ and the third component of $\mathbf{X}$ parallel to $\mathbf{c}^*$ and is

$$\boldsymbol{M} = \begin{pmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & c(\cos\alpha - \cos\beta\cos\gamma)/\sin\gamma \\ 0 & 0 & c\varphi/\sin\gamma \end{pmatrix}$$

$$\boldsymbol{M}^{-1} = \begin{pmatrix} 1/a & -1/a\tan\gamma & (\cos\alpha\cos\gamma - \cos\beta)/a\varphi\sin\gamma \\ 0 & 1/b\sin\gamma & (\cos\beta\cos\gamma - \cos\alpha)/b\varphi\sin\gamma \\ 0 & 0 & \sin\gamma/c\varphi \end{pmatrix}.$$

A third form, suitable only for rhombohedral cells, is

$$\boldsymbol{M} = \frac{a}{3}\begin{pmatrix} p+2q & p-q & p-q \\ p-q & p+2q & p-q \\ p-q & p-q & p+2q \end{pmatrix}$$

$$\boldsymbol{M}^{-1} = \frac{1}{3a}\begin{pmatrix} \frac{1}{p}+\frac{2}{q} & \frac{1}{p}-\frac{1}{q} & \frac{1}{p}-\frac{1}{q} \\ \frac{1}{p}-\frac{1}{q} & \frac{1}{p}+\frac{2}{q} & \frac{1}{p}-\frac{1}{q} \\ \frac{1}{p}-\frac{1}{q} & \frac{1}{p}-\frac{1}{q} & \frac{1}{p}+\frac{2}{q} \end{pmatrix}$$

in which

$$p = \pm\sqrt{1 + 2\cos\alpha} \quad q = \pm\sqrt{1 - \cos\alpha},$$

which preserves the equivalence of axes. Here the chiralities of the Cartesian and crystallographic axes are the same if p is chosen positive, and different otherwise, and the two sets of axes coincide in projection along the triad if q is chosen positive and are π out of phase otherwise.

3.3.1.1.2. Homogeneous coordinates

Homogeneous coordinates have found wide application in computer graphics. For some equipment their use is essential, and they are of value analytically even if the available hardware does not require their use.

Homogeneous coordinates employ four quantities, X, Y, Z and W, to define the position of a point, rather than three. The fourth coordinate has a scaling function so that it is the quantity X/W (as delivered to the display hardware) which controls the left–right positioning of the point within the picture. A point with $|X/W| < 1$ is in the picture, normally, and those with $|X/W| > 1$ are outside it, but see Section 3.3.1.3.5.

There are many reasons why homogeneous coordinates may be adopted, among them the following:

(i) X, Y, Z and W may be held as integers, thus enabling fast arithmetic whilst offering much of the flexibility of floating-point working. A single W value may be common to a whole array of X, Y, Z values.

(ii) Perspective transformations can be implemented without the need for any division. Only high-speed matrix multiplication using integer arithmetic is necessary, provided only that the drawing hardware can provide displacements proportional to the ratio of two signals, X and W or Y and W. Rotation, translation, scaling and the application of perspective are all effected by operations of the same form, namely multiplication of a four-vector by a 4×4 matrix. The hardware may thus be kept relatively simple since only one type of operation needs to be provided for.

(iii) Since kX, kY, kZ, kW represents the same point as X, Y, Z, W, the hardware may be arranged to maximize resolution without risk of integer overflow.

For analytical purposes it is convenient to regard homogeneous transformations in terms of partitioned matrices

$$\begin{pmatrix} \boldsymbol{M} & \mathbf{V} \\ \mathbf{U} & N \end{pmatrix}\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix},$$

Copyright © 2010 International Union of Crystallography

where $\boldsymbol{M}$ is a 3×3 matrix, $\mathbf{V}$ and $\mathbf{X}$ are three-element column vectors, $\mathbf{U}$ is a three-element row vector and N and W are scalars.

Matrices and vectors which are equivalent under the considerations of (iii) above will be related by the sign $\simeq$ in what follows.

Hardware systems which use true floating-point representations have less need of homogeneous coordinates and for these N and W may normally be set to unity.

3.3.1.1.3. Notation

In this chapter the conventions of matrix algebra will be adhered to except where it is convenient to show operations on elements of vectors, matrices and tensors, where a subscript notation will be used with a modified summation convention in which summation is over lower-case subscripts *only*. Thus the equation

$$x_I = A_{Ij}X_j$$

is to be read 'For any I, x_I is $A_{Ij}X_j$ summed over j'.

Subscripts using the letter i or later in the alphabet will relate to the usual three dimensions and imply a three-term summation. Subscripts a to h are not necessarily so limited, and, in particular, the subscript a is used to imply summation over atoms of which there may be an arbitrary number.

We shall use the superscript T to denote a transpose, and also use the Kronecker delta, δ_{IJ}, which is 1 if $I = J$ and zero otherwise, and the tensor ε_{IJK} which is 1 if I, J and K are a cyclic permutation of 1, 2, 3, -1 if an anticyclic permutation, and zero otherwise.

$$\varepsilon_{IJK} = (I - J)(J - K)(K - I)/2 \quad 1 \leq I, J, K \leq 3.$$

A useful identity is then

$$\varepsilon_{iJK}\varepsilon_{iLM} = \delta_{JL}\delta_{KM} - \delta_{JM}\delta_{KL}.$$

Single modulus signs surrounding the symbol for a square matrix denote its determinant, and around a vector denote its length.

The symbol $\simeq$ is defined in the previous section.

3.3.1.1.4. Standards

The sections of this chapter concerned with graphics are primarily concerned with the mathematical aspects of graphics programming as they confront the applications programmer. The implementations outlined in Section 3.3.3 have all, so far as the author is aware, been developed *ab initio* by their inventors to deal with these aspects using their own and unrelated techniques and protocols. It is clear, however, that standards are now emerging, and it is to be hoped that future developments in applications software will handle the graphics aspects through one or other of these standards.

First among these standards is the Graphical Kernel System, GKS, defined in *American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Graphical Kernel System (GKS) Functional Description* (1985) and described and illustrated by Hopgood *et al.* (1986) and Enderle *et al.* (1984). GKS became a full International Standards Organization (ISO) standard in 1985, and its purpose is to standardize the interface between application software and the graphics system, thus enhancing portability of software. Specifications for Fortran, Pascal and Ada formulations are at an advanced stage of development. Its value to crystallographers is limited by the fact that it is only two-dimensional. A three-dimensional extension known as GKS-3D, defined in *International Standards Organisation, International Standard Information Processing Systems – Computer Graphics – Graphical Kernel System for Three Dimensions (GKS-3D), Functional Description* (1988) became an ISO standard in 1988. Perhaps of greatest interest to crystallographers, however, is the Programmers' Hierarchical Interactive Graphics System (PHIGS) (Brown, 1985; Abi-Ezzi & Bunshaft, 1986) since this allows hierarchical segmentation of picture content to exist in both the applications software and the graphics device in a related manner, which GKS does not. Some graphics devices now available support this type of working and its exploitation indicates the choice of PHIGS. Furthermore, Fortran implementations of GKS and GKS-3D require points to be stored in arrays dimensioned as $X(N)$, $Y(N)$, $Z(N)$ which may be equivalenced (in the Fortran sense) to $XYZ(N, 3)$ but not to $XYZ(3, N)$, which may not be convenient. PHIGS also became an International Standard in 1988: *American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Programmer's Hierarchical Graphics System (PHIGS) Functional Description, Archive File Format, Clear-Text Encoding of Archive File* (1988). PHIGS has also been extended to support the capability of raster-graphics machines to represent reflections, shadows, see-through effects *etc.* in a version known as PHIGS+ (van Dam, 1988).

Increasingly, manufacturers of graphics equipment are orienting their products towards one or other of these standards. While these standards are not the subject of this chapter it is recommended that they be studied before investing in equipment.

In addition to these standards, related standards are evolving under the auspices of the ISO for defining images in a file-storage, or metafile, form, and for the interface between the device-independent and device-dependent parts of a graphics package. Arnold & Bono (1988) describe the ANSI and ISO Computer Graphics Metafile standard which provides for the definition of (two-dimensional) images. The definition of three-dimensional scenes requires the use of (PHIGS) archive files.

3.3.1.2. Orthogonal (or rotation) matrices

It is a basic requirement for any graphics or molecular-modelling system to be able to control and manipulate the orientation of the structures involved and this is achieved using orthogonal matrices which are the subject of these sections.

3.3.1.2.1. General form

If a vector $\mathbf{v}$ is expressed in terms of its components resolved onto an axial set of vectors $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ which are of unit length and mutually perpendicular and right handed in the sense that $(\mathbf{X} \times \mathbf{Y}) \cdot \mathbf{Z} = +1$, and if these components are v_I, and if a second set of axes $\mathbf{X}'$, $\mathbf{Y}'$, $\mathbf{Z}'$ is similarly established, with the same origin and chirality, and if $\mathbf{v}$ has components v'_I on these axes then

$$v'_I = a_{Ij}v_j,$$

in which a_{IJ} is the cosine of the angle between the ith primed axis and the jth unprimed axis. Evidently the elements a_{IJ} comprise a matrix $\boldsymbol{R}$, such that any row represents one of the primed axial vectors, such as $\mathbf{X}'$, expressed as components on the unprimed axes, and each column represents one of the unprimed axial vectors expressed as components on the primed axes. It follows that $\boldsymbol{R}^T = \boldsymbol{R}^{-1}$ since elements of the product $\boldsymbol{R}^T\boldsymbol{R}$ are scalar products among perpendicular unit vectors.

A real matrix whose transpose equals its inverse is said to be *orthogonal.*

Since $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ can simultaneously be superimposed on $\mathbf{X}'$, $\mathbf{Y}'$ and $\mathbf{Z}'$ without deformation or change of scale the relationship is one of rotation, and orthogonal matrices are often referred to as rotation matrices. The operation of replacing the vector $\mathbf{v}$ by $\boldsymbol{R}\mathbf{v}$ corresponds to rotating the axes from the unprimed to the primed set with $\mathbf{v}$ itself unchanged. Equally, the same operation corresponds to retaining fixed axes and rotating the vector in the opposite sense. The second interpretation is the one more frequently helpful since conceptually it corresponds more closely

to rotational operations on objects, and it is primarily in this sense that the following is written.

If three vectors **u**, **v** and **w** form the edges of a parallelepiped, then its volume V is

$$V = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \varepsilon_{ijk} u_i v_j w_k$$

and if these vectors are transformed by the matrix $\boldsymbol{R}$ as above, then the transformed volume V' is

$$V' = \varepsilon_{lmn} u'_l v'_m w'_n = \varepsilon_{lmn} a_{li} a_{mj} a_{nk} u_i v_j w_k.$$

But the determinant of $\boldsymbol{R}$ is given by

$$|\boldsymbol{R}| \varepsilon_{IJK} = \varepsilon_{lmn} a_{lI} a_{mJ} a_{nK}$$

so that

$$V' = |\boldsymbol{R}| V$$

and the determinant of $\boldsymbol{R}$ must therefore be +1 for a transformation which is a pure rotation. Nevertheless orthogonal matrices with determinant -1 exist though these do not describe a pure rotation. They may always be described as the product of a pure rotation and inversion through the origin and are referred to here as improper rotations. In what follows all references to orthogonal matrices refer to those with positive determinant only, unless stated otherwise.

An important general form of an orthogonal matrix in three dimensions was derived as equation (1.1.4.32) and is

$$\boldsymbol{R} = \begin{pmatrix} l^2 + (m^2 + n^2)\cos\theta & lm(1 - \cos\theta) - n\sin\theta & nl(1 - \cos\theta) + m\sin\theta \\ lm(1 - \cos\theta) + n\sin\theta & m^2 + (n^2 + l^2)\cos\theta & mn(1 - \cos\theta) - l\sin\theta \\ nl(1 - \cos\theta) - m\sin\theta & mn(1 - \cos\theta) + l\sin\theta & n^2 + (l^2 + m^2)\cos\theta \end{pmatrix}$$

or

$$R_{IJ} = (1 - \cos\theta) l_I l_J + \delta_{IJ} \cos\theta - \varepsilon_{IJk} l_k \sin\theta,$$

in which l, m and n are the direction cosines of the axis of rotation (which are the same when referred to either set of axes under either interpretation) and θ is the angle of rotation. In this form, and with $\boldsymbol{R}$ operating on column vectors on the right, the sign of θ is such that, when viewed along the rotation axis *from* the origin *towards* the point *lmn,* the object is rotated clockwise for positive θ with a fixed right-handed axial system. If, under the same viewing conditions, the axes are to be rotated clockwise through θ with the object fixed then the components of vectors in the object, on the new axes, are given by $\boldsymbol{R}$ with the same *lmn* and with θ negated. This is the transpose of $\boldsymbol{R}$, and if $\boldsymbol{R}$ is constructed from a product, as below, then each factor matrix in the product must be transposed and their order reversed to achieve this. Note that if, for a given rotation, the viewing direction from the origin is reversed, l, m, n and θ are all reversed and the matrix is unchanged.

Any rotation about a reference axis such that two of the direction cosines are zero is termed a *primitive rotation*, and it is frequently a requirement to generate or to interpret a general rotation as a product of primitive rotations.

A second important general form is based on Eulerian angles and is the product of three such primitives. It is

$$\begin{aligned} \boldsymbol{R} &= \begin{pmatrix} \cos\varphi_3 & -\sin\varphi_3 & 0 \\ \sin\varphi_3 & \cos\varphi_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\varphi_2 & 0 & \sin\varphi_2 \\ 0 & 1 & 0 \\ -\sin\varphi_2 & 0 & \cos\varphi_2 \end{pmatrix} \begin{pmatrix} \cos\varphi_1 & -\sin\varphi_1 & 0 \\ \sin\varphi_1 & \cos\varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} (\cos\varphi_3\cos\varphi_2\cos\varphi_1 & -(\cos\varphi_3\cos\varphi_2\sin\varphi_1 & \cos\varphi_3\sin\varphi_2 \\ \quad -\sin\varphi_3\sin\varphi_1) & \quad +\sin\varphi_3\cos\varphi_1) & \\ (\sin\varphi_3\cos\varphi_2\cos\varphi_1 & (-\sin\varphi_3\cos\varphi_2\sin\varphi_1 & \sin\varphi_3\sin\varphi_2 \\ \quad +\cos\varphi_3\sin\varphi_1) & \quad +\cos\varphi_3\cos\varphi_1) & \\ -\sin\varphi_2\cos\varphi_1 & \sin\varphi_2\sin\varphi_1 & \cos\varphi_2 \end{pmatrix} \end{aligned}$$

which is commonly employed in four-circle diffractometers for which $\varphi = -\varphi_1$, $\chi = \varphi_2$ and $\omega = -\varphi_3$. In terms of the fixed-axes–moving-object conceptualization this corresponds to a rotation φ_1 about Z followed by φ_2 about Y followed by φ_3 about Z. In the familiar diffractometer example, when $\chi = 0$ the φ and ω axes are both vertical and equivalent. If φ is altered first, then the χ axis is still in the direction of a fixed Y axis, but if ω is altered first it is not. Since all angles are to be rotations about fixed axes to describe a rotating object it follows that it is φ rather than ω which corresponds to φ_1. In general, when rotating parts are mounted on rotating parts the rotation closest to the moved object must be applied first, forming the right-most factor in any multiple transformation, with the rotation closest to the fixed part as the left-most factor, assuming data supplied as column vectors on the right.

Given an orthogonal matrix, in either numerical or analytical form, it may be required to discover θ and the axis of rotation, or to factorize it as a product of primitives. From the first form we see that the vector

$$v_I = \varepsilon_{Ijk} a_{jk},$$

consisting of the antisymmetric part of $\boldsymbol{R}$, has elements $-2\sin\theta$ times the direction cosines l, m, n, which establishes the direction immediately, and normalization using $l^2 + m^2 + n^2 = 1$ determines $\sin\theta$. Furthermore, the trace is $1 + 2\cos\theta$ so that the quadrant of θ is also fixed. This method fails, however, if the matrix is symmetrical, which occurs if $\theta = \pi$. In this case only the direction of the axis is required, which is given by

$$l : m : n = (a_{23})^{-1} : (a_{31})^{-1} : (a_{12})^{-1}$$

for nonzero elements, or $l = \sqrt{\frac{1}{2}(a_{11} + 1)}$ *etc.*, with the signs chosen to satisfy $a_{12} = 2lm$ *etc.*

The Eulerian form may be factorized by noting that $\tan\varphi_1 = -a_{32}/a_{31}$, $\tan\varphi_3 = a_{23}/a_{13}$, $\cos\varphi_2 = a_{33}$. There is then freedom to choose the sign of $\sin\varphi_2$, but the choice then fixes the quadrants of φ_1 and φ_3 through the elements in the last row and column, and the primitives may then be constructed. These expressions for φ_1 and φ_3 fail if $\sin\varphi_2 = 0$, in which case the rotation reduces to a primitive rotation about Z with angle $(\varphi_1 + \varphi_3)$, $\varphi_2 = 0$, or $(\varphi_3 - \varphi_1)$, $\varphi_2 = \pi$.

Eulerian angles are unlikely to be the best choice of primitive angles unless they are directly related to the parameters of a system, as with the diffractometer. It is often more important that the changes to primitive angles should be quasi-linearly related to θ for any small rotations, which is not the case with Eulerian angles when the required rotation axis is close to the X axis. In such a case linearized techniques for solving for the primitive angles will fail. Furthermore, if the required rotation is about Z only $(\varphi_1 + \varphi_3)$ is determinate.

Quasi-linear relationships between θ and the primitive rotations arise if the primitives are one each about X, Y and Z. Any order of the three factors may be chosen, but the choice must then be adhered to since these factors do not commute. For sufficiently small rotations the primitive rotations are then $l\theta$, $m\theta$ and $n\theta$, whilst for larger θ linearized iterative techniques for finding the primitive rotations are likely to be convergent and well conditioned.

The three-dimensional space of the angles φ_1, φ_2 and φ_3 in either case is nonlinearly related to θ. In the Eulerian case the worst nonlinearities occur at the origin of φ-space. Equally severe nonlinearities occur in the quasi-linear case also but are 90° away from the origin and less likely to be troublesome.

Neither of the foregoing general forms of orthogonal matrix has ideally convenient properties. The first is inconvenient because it uses four nonequivalent variables l, m, n and θ, with a linking equation involving l, m and n, so that they cannot be treated as independent variables for analytical purposes. The

second form (the product of primitives) is not ideal because the three angles, though independent, are not equivalent, the non-equivalence arising from the noncommutation of the primitive factors. In the remainder of this section we give two further forms of orthogonal matrix which each use three variables which are independent and strictly equivalent, and a third form using four whose squares sum to unity.

The first of these is based on the diagonal and uses the three independent variables p, q, r, from which we construct the auxiliary variables

$$P = \pm\sqrt{1+p-q-r},\ \ Q = \pm\sqrt{1-p+q-r},$$
$$R = \pm\sqrt{1-p-q+r},\ \ S = \pm\sqrt{1+p+q+r},$$

then

$$\boldsymbol{R} = \begin{pmatrix} p & \frac{1}{2}[PQ-RS] & \frac{1}{2}[PR+QS] \\ \frac{1}{2}[PQ+RS] & q & \frac{1}{2}[QR-PS] \\ \frac{1}{2}[PR-QS] & \frac{1}{2}[QR+PS] & r \end{pmatrix}$$

is orthogonal with positive determinant for any of the sixteen sign combinations. The signs of P, Q, R and S are, respectively, the signs of the direction cosines of the rotation axis and of $\sin\theta$. Using also $T = \sqrt{4-S^2}$, which may be deemed positive without loss of generality,

$$l = P/T, m = Q/T, n = R/T, \sin\theta = ST/2,$$
$$\cos\theta = 1 - T^2/2 = S^2/2 - 1.$$

Although p, q and r are independent, the point $[pqr]$ is bound, by the requirement that P, Q, R and S be real, to lie within a tetrahedron whose vertices are the points [111], $[1\bar{1}\bar{1}]$, $[\bar{1}1\bar{1}]$ and $[\bar{1}\bar{1}1]$, corresponding to the identity and to 180° rotations about each of the axes. The facts that the identity occurs at a vertex of the feasible region and that $(1-\cos\theta)$, rather than $\sin\theta$, is linear on p, q and r in this vicinity make this form suitable only for substantial rotations.

The second form consists in defining a rotation vector $\mathbf{r}$ with components u, v, w such that $u = lt$, $v = mt$, $w = nt$ with $t = \tan(\theta/2)$ and $\mathbf{r}\cdot\mathbf{r} = t^2$. Then the matrix

$$\boldsymbol{R} = \begin{pmatrix} \dfrac{1+u^2-v^2-w^2}{1+t^2} & \dfrac{2(uv-w)}{1+t^2} & \dfrac{2(uw+v)}{1+t^2} \\ \dfrac{2(uv+w)}{1+t^2} & \dfrac{1-u^2+v^2-w^2}{1+t^2} & \dfrac{2(vw-u)}{1+t^2} \\ \dfrac{2(uw-v)}{1+t^2} & \dfrac{2(vw+u)}{1+t^2} & \dfrac{1-u^2-v^2+w^2}{1+t^2} \end{pmatrix}$$

$$R_{IJ} = (1+t^2)^{-1}[\delta_{IJ}(1-u_k u_k) + 2(u_I u_J - \varepsilon_{IJl}u_l)]$$

is orthogonal and the variables u, v, w are independent, equivalent and unbounded, and, unlike the previous form, small rotations are quasi-linear on these variables. As examples, $\mathbf{r}$ = [100] gives 90° about X, $\mathbf{r}$ = [111] gives 120° about [111].

$\boldsymbol{R}$ then transforms a vector $\mathbf{d}$ according to

$$\boldsymbol{R}\mathbf{d} = \mathbf{d} + \frac{2}{1+t^2}\{(\mathbf{r}\times\mathbf{d}) + [\mathbf{r}\times(\mathbf{r}\times\mathbf{d})]\}.$$

Multiplying two such matrices together allows us to establish the manner in which the rotation vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ combine.

$$\mathbf{r} = \frac{\mathbf{r}_2 + \mathbf{r}_1 + \mathbf{r}_2\times\mathbf{r}_1}{1-\mathbf{r}_2\cdot\mathbf{r}_1}$$

for a rotation $\mathbf{r}_1$ followed by $\mathbf{r}_2$, so that rotations expressed in terms of rotation angles and axes may be compounded into a single such rotation without the need to form and decompose a product matrix.

Note that if $\mathbf{r}_1$ and $\mathbf{r}_2$ are parallel this reduces to the formula for the tangent of the sum of two angles, and that if $\mathbf{r}_1\cdot\mathbf{r}_2 = 1$ the combined rotation is always 180°. Note, too, that reversing the order of application of the rotations reverses only the vector product.

If three rotations $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are applied successively, $\mathbf{r}_1$ first, then their combined rotation is

$$\begin{aligned}\mathbf{r} = [&\mathbf{r}_3(1-\mathbf{r}_1\cdot\mathbf{r}_2) + \mathbf{r}_2(1+\mathbf{r}_3\cdot\mathbf{r}_1) + \mathbf{r}_1(1-\mathbf{r}_3\cdot\mathbf{r}_2) \\ &+ \mathbf{r}_3\times\mathbf{r}_2 + \mathbf{r}_3\times\mathbf{r}_1 + \mathbf{r}_2\times\mathbf{r}_1] \\ &\times[1-\mathbf{r}_1\cdot\mathbf{r}_2 - \mathbf{r}_2\cdot\mathbf{r}_3 - \mathbf{r}_3\cdot\mathbf{r}_1 - \mathbf{r}_3\cdot(\mathbf{r}_2\times\mathbf{r}_1)]^{-1}.\end{aligned}$$

Note the irregular pattern of signs in the numerator.

Similar ideas, using a vector of magnitude $\sin(\theta/2)$, are developed in Aharonov *et al.* (1977).

The third form of orthogonal matrix uses four variables, λ, μ, ν and σ, which comprise a four-dimensional vector $\boldsymbol{\rho}$, such that $\lambda = ls$, $\mu = ms$, $\nu = ns$ with $s = \sin(\theta/2)$ and $\sigma = \cos(\theta/2)$. In terms of these variables

$$\boldsymbol{R} = \begin{pmatrix} (\lambda^2-\mu^2-\nu^2+\sigma^2) & 2(\lambda\mu-\nu\sigma) & 2(\lambda\nu+\mu\sigma) \\ 2(\mu\lambda+\nu\sigma) & (-\lambda^2+\mu^2-\nu^2+\sigma^2) & 2(\mu\nu-\lambda\sigma) \\ 2(\lambda\nu-\mu\sigma) & 2(\mu\nu+\lambda\sigma) & (-\lambda^2-\mu^2+\nu^2+\sigma^2) \end{pmatrix}.$$

Two further matrices $\boldsymbol{S}$ and $\boldsymbol{T}$ may be defined (Diamond, 1988),

$$\boldsymbol{S} = \begin{pmatrix} -\sigma & \nu & -\mu & \lambda \\ -\nu & -\sigma & \lambda & \mu \\ \mu & -\lambda & -\sigma & \nu \\ \lambda & \mu & \nu & \sigma \end{pmatrix} \text{ and } \boldsymbol{T} = \begin{pmatrix} \sigma & -\nu & \mu & \lambda \\ \nu & \sigma & -\lambda & \mu \\ -\mu & \lambda & \sigma & \nu \\ -\lambda & -\mu & -\nu & \sigma \end{pmatrix},$$

which are themselves orthogonal (though $\boldsymbol{S}$ has determinant -1) and which have the property that

$$\boldsymbol{S}^2 = \begin{pmatrix} \boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}$$

so that, for example, if homogeneous coordinates are being employed (Section 3.3.1.1.2)

$$\begin{pmatrix} x' \\ y' \\ z' \\ w \end{pmatrix} = \begin{pmatrix} -\sigma & \nu & -\mu & \lambda \\ -\nu & -\sigma & \lambda & \mu \\ \mu & -\lambda & -\sigma & \nu \\ \lambda & \mu & \nu & \sigma \end{pmatrix}\begin{pmatrix} -\sigma & \nu & -\mu & \lambda \\ -\nu & -\sigma & \lambda & \mu \\ \mu & -\lambda & -\sigma & \nu \\ \lambda & \mu & \nu & \sigma \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$$

is a rotation of (x, y, z, w) through the angle θ about the axis (l, m, n). With suitably pipelined hardware this forms an efficient means of applying rotations since the 'overhead' of establishing $\boldsymbol{S}$ is so trivial.

$\boldsymbol{T}$ has the property that the rotation vector $\boldsymbol{\rho}$ arising from a concatenation of n rotations is

$$\boldsymbol{\rho} = \boldsymbol{T}_n\boldsymbol{T}_{n-1}\ldots\boldsymbol{T}_1\boldsymbol{\rho}_0,$$

in which $\boldsymbol{\rho}_0^T$ is the vector (0, 0, 0, 1) which defines a null rotation. This equation may be used as a basis for factorizing a given rotation into a concatenation of rotations about designated axes (Diamond, 1990*a*).

Finally, an exact rotation of the vector $\mathbf{d}$ may be obtained without using matrices at all by writing

$$\mathbf{d} = \sum_0^\infty \mathbf{d}_n$$

in which

$$\mathbf{d}_n = \frac{1}{n}(\boldsymbol{\theta}\times\mathbf{d}_{n-1})$$

and $\mathbf{d}_0$ is the initial position which is to be rotated. Here $\boldsymbol{\theta}$ is a vector with direction cosines l, m and n, and magnitude equal to

the required rotation angle in radians (Diamond, 1966). This method is particularly efficient when $|\boldsymbol{\theta}| \ll 1$ or when the number of vectors to be transformed is small since the overhead of establishing $\boldsymbol{R}$ is eliminated and the process is simple to program. It is the three-dimensional analogue of the power series for $\sin\theta$ and $\cos\theta$ and has the same convergence properties.

3.3.1.2.2. Measurement of rotations and strains from coordinates

Given the coordinates of a molecular fragment it is often a requirement to relate the fragment to its image in some standard orientation by a transformation which may be required to be a pure rotation, or may be required to be a combination of rotation and strain. Of the methods reviewed in this section all except (iv) are concerned with pure rotation, ignoring any strain that may be present, and give the best rigid-body superposition. In all these methods, unless inhomogeneous strain is being considered, the best possible superposition is obtained if the centroids of the two images are first brought into coincidence by translation and treated as the origin.

Methods (i) to (v) seek transformations which perform the superposition and impose on these, in various ways, the requirements of orthogonality for the rotational part. All these methods therefore need some defence against indeterminacy that arises in the general transformation if one or both of the fragments is planar, and, if improper rotations are to be excluded, need a defence against these also if the fragment and its image are of opposite chirality. Methods (vi) and (vii) pay no attention to the general transformation (which defines the general superposition) and work with variables which are intrinsically rotational in character, and always produce an orthogonal transformation with positive determinant, with no degeneracy arising from planar fragments which need no special attention. Even collinear atoms cause no problem, the superposition being performed correctly but with an arbitrary rotation about the length of the line being present in the result. These methods are therefore to be preferred over the earlier ones unless the purpose of the operation is to detect differences of chirality, although this, too, can be detected with a simple test.

In this review we adopt the same notation for all the methods which, unavoidably, means that symbols are used in ways which differ from the original publications. We use the symbol $\boldsymbol{x}$ for the vector set which is to be rotated and $\boldsymbol{X}$ for the vector set whose orientation is not to be altered, and write the residuals as

$$e_{IA} = D_{Ij}x_{jA} - X_{IA}$$

and, by choice of origin,

$$W_a x_{Ia} = W_a X_{Ia} = 0_I$$

for weights W. The quadratic residual to be minimized is

$$E = W_a e_{ia} e_{ia}$$

and we define the matrix $M_{IJ} = W_a x_{Ia} X_{Ja}$ and use $\mathbf{l}$ for the direction cosines of the rotation axis.

(i) McLachlan's first method (McLachlan, 1972, 1982) is iterative and conceptually the simplest. It sets

$$D_{IJ} = A_{Ik}R_{kJ}$$

in which $\boldsymbol{A}$ and $\boldsymbol{R}$ are both orthogonal with $\boldsymbol{R}$ being a current estimate of $\boldsymbol{D}$ and $\boldsymbol{A}$ being an adjustment which, at the beginning of each cycle, has a zero angle associated with it. One iterative cycle estimates a nontrivial $\boldsymbol{A}$, after which the product $\boldsymbol{AR}$ replaces $\boldsymbol{R}$.

$$A_{IJ} = (1 - \cos\theta)l_I l_J + \delta_{IJ}\cos\theta - \varepsilon_{IJk}l_k \sin\theta$$

and

$$\left(\frac{\partial A_{IJ}}{\partial\theta}\right)_{\theta=0} = -\varepsilon_{IJk}l_k,$$

therefore

$$\left(\frac{\partial E}{\partial\theta}\right)_{\theta=0} = 2W_a\left(\frac{\partial A_{ij}}{\partial\theta}\right)_{\theta=0} R_{jk}x_{ka}(A_{il}R_{lm}x_{ma} - X_{ia})$$
$$= 2\varepsilon_{ijl}R_{jk}M_{ki}l_l.$$

For this to vanish for all possible rotation axes $\mathbf{l}$ the vector

$$g_L = \varepsilon_{ijL}R_{jk}M_{ki}$$

must vanish, *i.e.* at the end of the iteration $\boldsymbol{R}$ must be such that the matrix

$$N_{JI} = R_{Jk}M_{kI}$$

is symmetrical. The vector $\mathbf{g}$ represents the couple exerted on the rotating body by forces $2W_A(R_{Ij}x_{jA} - X_{IA})$ acting at the atoms. Choosing

$$l_L = g_L/|\mathbf{g}|$$

gives the greatest $|\partial E/\partial\theta|_{\theta=0}$ and $(\partial E/\partial\theta)$ vanishes when

$$\tan\theta = \frac{\varepsilon_{ijk}N_{ji}l_k}{N_{pq}(l_p l_q - \delta_{pq})}$$

in which $\boldsymbol{N}$ is constructed from the current $\boldsymbol{R}$ matrix. $\boldsymbol{A}$ is then constructed from $\mathbf{l}$ and this θ and $\boldsymbol{AR}$ replaces $\boldsymbol{R}$. The process is iterative because a couple about some new axis can appear when rotation about $\mathbf{g}$ eliminates the couple about $\mathbf{g}$.

Note that for each rotation axis $\mathbf{l}$ there are two values of θ, differing by π, which reduce $|\mathbf{g}|$ to zero, corresponding to maximum and minimum values of E. The minimum is that which makes

$$\frac{\partial^2 E}{\partial\theta^2} = 2(\text{tr } N - l_i N_{ij} l_j)$$

positive. Adding π to θ alters $\boldsymbol{R}$ and $\boldsymbol{N}$ and negates this quantity.

Note, too, that the process is essentially characterized as that which makes the product $\boldsymbol{RM}$ symmetrical with $\boldsymbol{R}$ orthogonal. We return to this point in (iii).

(ii) Kabsch's method (Kabsch, 1976, 1978) minimizes E with respect to the nine elements of $\boldsymbol{D}$, subject to the six constraints

$$D_{kI}D_{kJ} - \delta_{IJ} = 0_{IJ},$$

by using an auxiliary function

$$F = L_{ij}(D_{ki}D_{kj} - \delta_{ij})$$

in which $\boldsymbol{L}$ is symmetric containing six Lagrange multipliers. The Lagrangian function

$$G = E + F$$

then has minima with respect to the elements of $\boldsymbol{D}$ at locations which are dependent, *inter alia*, on the elements of $\boldsymbol{L}$. By suitably choosing $\boldsymbol{L}$ a minimum of G may be brought into coincidence with the constrained minimum of $\boldsymbol{E}$. A minimum of G occurs where

$$\frac{\partial G}{\partial D_{IJ}} = 2D_{Ik}(S_{Jk} + L_{Jk}) - 2M_{JI} = 0_{IJ}$$

and the 9×9 matrix

$$\frac{\partial^2 G}{\partial D_{MK}\partial D_{IJ}} = 2\delta_{MI}(S_{JK} + L_{JK})$$

is positive definite, block diagonal, and has

$$S_{JK} = W_a x_{Ja} x_{Ka}$$

which is symmetrical. Thus $\boldsymbol{L}$ must be chosen so as to make the symmetric matrix $(\boldsymbol{S}+\boldsymbol{L})$ such that

$$\boldsymbol{D}(\boldsymbol{S}+\boldsymbol{L})^T = \boldsymbol{M}^T$$

with $\boldsymbol{D}$ orthogonal, or $\boldsymbol{RN} = \boldsymbol{M}^T$ with $\boldsymbol{R}$ replacing $\boldsymbol{D}$ since we are now confined to the orthogonal case, and $\boldsymbol{N}$ is symmetric and positive definite.

(iii) Comparison of the Kabsch and McLachlan methods. Using the initials of these authors as subscripts, we have seen that the Kabsch solution involves solving

$$\boldsymbol{R}_{\rm WK}\boldsymbol{N}_{\rm WK} = \boldsymbol{M}^T$$

for an orthogonal matrix $\boldsymbol{R}_{\rm WK}$ given that $\boldsymbol{N}_{\rm WK}$ is symmetrical and positive definite. Thus

$$\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{N}_{\rm WK}^T\boldsymbol{R}_{\rm WK}^T\boldsymbol{R}_{\rm WK}\boldsymbol{N}_{\rm WK} = \boldsymbol{N}_{\rm WK}^2$$

and

$$\boldsymbol{R}_{\rm WK} = \boldsymbol{M}^T(\boldsymbol{M}\boldsymbol{M}^T)^{-1/2}.$$

By comparison, the McLachlan treatment leads to an orthogonal $\boldsymbol{R}$ matrix satisfying

$$\boldsymbol{R}_{\rm ADM} = \boldsymbol{N}_{\rm ADM}\boldsymbol{M}^{-1}$$

in which $\boldsymbol{N}_{\rm ADM}$ is also symmetric and positive definite, which similarly leads to

$$\boldsymbol{R}_{\rm ADM} = (\boldsymbol{M}^T\boldsymbol{M})^{1/2}\boldsymbol{M}^{-1}.$$

These seemingly different expressions for $\boldsymbol{R}_{\rm WK}$ and $\boldsymbol{R}_{\rm ADM}$ are, in fact, equal, as the following shows

$$\boldsymbol{R}_{\rm WK} = \boldsymbol{R}_{\rm ADM}\boldsymbol{R}_{\rm ADM}^{-1}\boldsymbol{R}_{\rm WK} = \boldsymbol{R}_{\rm ADM}\boldsymbol{M}\boldsymbol{N}_{\rm ADM}^{-1}\boldsymbol{M}^T\boldsymbol{N}_{\rm WK}^{-1},$$

therefore

$$\begin{aligned}\boldsymbol{R}_{\rm WK}^T\boldsymbol{R}_{\rm WK} &= \boldsymbol{I}\\ &= \boldsymbol{N}_{\rm WK}^{-1}\boldsymbol{M}\boldsymbol{N}_{\rm ADM}^{-1}\boldsymbol{M}^T\boldsymbol{R}_{\rm ADM}^T\boldsymbol{R}_{\rm ADM}\boldsymbol{M}\boldsymbol{N}_{\rm ADM}^{-1}\boldsymbol{M}^T\boldsymbol{N}_{\rm WK}^{-1}.\end{aligned}$$

Multiplying on both sides by $\boldsymbol{N}_{\rm WK}$ gives

$$\boldsymbol{N}_{\rm WK}^2 = (\boldsymbol{M}\boldsymbol{N}_{\rm ADM}^{-1}\boldsymbol{M}^T)^2,$$

and since both $\boldsymbol{N}$ matrices are positive definite

$$\boldsymbol{N}_{\rm WK} = \boldsymbol{M}\boldsymbol{N}_{\rm ADM}^{-1}\boldsymbol{M}^T$$

and conversely

$$\boldsymbol{N}_{\rm ADM} = \boldsymbol{M}^T\boldsymbol{N}_{\rm WK}^{-1}\boldsymbol{M},$$

therefore

$$\boldsymbol{R}_{\rm WK} = \boldsymbol{M}^T\boldsymbol{M}^{T-1}\boldsymbol{N}_{\rm ADM}\boldsymbol{M}^{-1} = \boldsymbol{R}_{\rm ADM}.$$

(iv) Diamond's first method. This method (Diamond, 1976*a*) differs from the previous ones in that the transformation $\boldsymbol{D}$ is allowed to be a general transformation which is then factorized into the product of an orthogonal matrix $\boldsymbol{R}$ and a symmetrical matrix $\boldsymbol{T}$. The transformation of $\boldsymbol{x}$ to fit $\boldsymbol{X}$ is thus interpreted as the combination of homogeneous strain and pure rotation in which $\boldsymbol{x}$ is subjected to strain and the result is rotated.

$$\begin{aligned}\boldsymbol{D} &= \boldsymbol{R}\boldsymbol{T}\\ \boldsymbol{T}^2 &= \boldsymbol{D}^T\boldsymbol{D}\\ \boldsymbol{T} &= (\boldsymbol{D}^T\boldsymbol{D})^{1/2}\\ \boldsymbol{R} &= \boldsymbol{D}(\boldsymbol{D}^T\boldsymbol{D})^{-1/2}.\end{aligned}$$

Furthermore, the solution for $\boldsymbol{D}$ is

$$\boldsymbol{D} = \boldsymbol{M}^T\boldsymbol{S}^{-1}$$

(in the notation of Kabsch), so that

$$\boldsymbol{R} = \boldsymbol{M}^T\boldsymbol{S}^{-1}(\boldsymbol{S}^{-1}\boldsymbol{M}\boldsymbol{M}^T\boldsymbol{S}^{-1})^{-1/2}$$

which may be compared with the results of the previous paragraph.

Although this $\boldsymbol{R}$ matrix by itself (*i.e.* applied without $\boldsymbol{T}$) does not produce the best rotational superposition (*i.e.* smallest E), it is the one which *exactly* superposes the only three vectors in $\boldsymbol{x}$ whose mutual dispositions are conserved, on their equivalents in $\boldsymbol{X}$, so that the rotation so found is arguably the best defined one.

Alternatives based on $\boldsymbol{D} = \boldsymbol{T}\boldsymbol{R}$, $\boldsymbol{D}^{-1} = \boldsymbol{R}\boldsymbol{T}$, $\boldsymbol{D}^{-1} = \boldsymbol{T}\boldsymbol{R}$ are all easily developed, and these ideas are developed by Diamond (1976*a*) to include nonhomogeneous strains also.

(v) McLachlan's second method. This method (McLachlan, 1979) is based on the properties of the 6×6 matrix

$$\begin{pmatrix}\boldsymbol{0} & \boldsymbol{M}\\ \boldsymbol{M}^T & \boldsymbol{0}\end{pmatrix}$$

and is immune to singularity of $\boldsymbol{M}$. If $\mathbf{p}$ and $\mathbf{q}$ are three-dimensional vectors such that $(\mathbf{p}^T, \mathbf{q}^T)$ is an eigenvector of this matrix then

$$\begin{pmatrix}\boldsymbol{0} & \boldsymbol{M}\\ \boldsymbol{M}^T & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\boldsymbol{M}\mathbf{q}\\ \boldsymbol{M}^T\mathbf{p}\end{pmatrix} = \begin{pmatrix}\mathbf{p}\lambda\\ \mathbf{q}\lambda\end{pmatrix}.$$

If $\mathbf{q}$ is negated the second equality is maintained provided λ is also negated. Therefore an orthogonal 6×6 matrix

$$\begin{pmatrix}\boldsymbol{H} & \boldsymbol{H}\\ \boldsymbol{K} & -\boldsymbol{K}\end{pmatrix}$$

(consisting of 3×3 partitions) exists for which

$$\begin{pmatrix}\boldsymbol{H}^T & \boldsymbol{K}^T\\ \boldsymbol{H}^T & -\boldsymbol{K}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{0} & \boldsymbol{M}\\ \boldsymbol{M}^T & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{H} & \boldsymbol{H}\\ \boldsymbol{K} & -\boldsymbol{K}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Lambda} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{\Lambda}\end{pmatrix}$$

in which $\boldsymbol{\Lambda}$ is diagonal and contains non-negative eigenvalues. The reverse transformation shows that

$$\boldsymbol{M} = 2\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{K}^T$$

and multiplying the eigenvectors together gives

$$\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{K}^T\boldsymbol{K} = \tfrac{1}{2}\boldsymbol{I} = \boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{K}\boldsymbol{K}^T.$$

Therefore

$$2\boldsymbol{K}\boldsymbol{H}^T\boldsymbol{M} = 4\boldsymbol{K}\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{K}^T = 2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{K}^T,$$

but $2\boldsymbol{K}\boldsymbol{H}^T$ is orthogonal and $2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{K}^T$ is symmetrical, therefore [by paragraphs (i) and (iii) above] $2\boldsymbol{K}\boldsymbol{H}^T$ is the required rotation. Similarly, forming

$$\begin{aligned}\boldsymbol{M}^T &= 2\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{H}^T\\ 2\boldsymbol{M}^T\boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^T &= 4\boldsymbol{K}\boldsymbol{\Lambda}\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^T = 2\boldsymbol{K}\boldsymbol{H}^T\end{aligned}$$

corresponds to the Kabsch formulation [paragraphs (ii) and (iii)] since $2\boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^T$ is symmetrical and the same rotation, $2\boldsymbol{K}\boldsymbol{H}^T$, appears.

Note that the determinant of the orthogonal matrix so found is twice the product of the determinants of $\boldsymbol{H}$ and of $\boldsymbol{K}$, and since the positive eigenvalues are collected into $\boldsymbol{\Lambda}$ it follows that the sign of the determinant of $\boldsymbol{M}$ is the same as the sign of the determinant of the resulting orthogonal matrix. If this is negative it means that the best superposition is obtained if one vector set is inverted and that $\boldsymbol{x}$ and $\boldsymbol{X}$ are of opposite chirality.

Expanding the expression for E, the weighted sum of squares of errors, for an orthogonal transformation shows that this is least when the trace of the product $\boldsymbol{RM}$ is greatest. In this treatment

$$\mathrm{tr}(\boldsymbol{RM}) = \mathrm{tr}(2\boldsymbol{KH}^T \cdot 2\boldsymbol{H\Lambda K}^T) = \mathrm{tr}(2\boldsymbol{K\Lambda K}^T) = \mathrm{tr}(\boldsymbol{\Lambda}).$$

Hence, if the eigenvalues in $\boldsymbol{\Lambda}$ and $-\boldsymbol{\Lambda}$ are arranged in decreasing order of modulus, and if the determinant of $\boldsymbol{M}$ is negative, then exchanging the third and sixth columns of

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{H} \\ \boldsymbol{K} & -\boldsymbol{K} \end{pmatrix}$$

produces a product $2\boldsymbol{KH}^T$ with positive determinant (*i.e.* a proper rotation) at minimum cost in residual. Similarly, if $\boldsymbol{M}$ is singular and one or more eigenvalues in $\boldsymbol{\Lambda}$ vanishes it is necessary only to complete an orthonormal set of eigenvectors such that the determinants of $\boldsymbol{H}$ and $\boldsymbol{K}$ have the same sign.

(vi) MacKay's method. MacKay (1984) was the first to consider the rotational superposition problem in terms of the vector $\mathbf{r}$ of Section 3.3.1.2.1. Using quaternion algebra he showed that if a vector $\mathbf{x}$ is rotated to $\mathbf{X} = \boldsymbol{R}\mathbf{x}$ then

$$(\mathbf{X} - \mathbf{x}) = \mathbf{r} \times (\mathbf{X} + \mathbf{x}),$$

where $|\mathbf{r}| = \tan(\theta/2)$ and the direction of $\mathbf{r}$ is the axis of rotation, as may also be shown from elementary considerations. MacKay then solves this for the vector $\mathbf{r}$ by least squares given the vector pairs $\mathbf{X}$ and $\mathbf{x}$. The individual errors are

$$e_{IA} = \varepsilon_{Ijk} r_j (X_{kA} + x_{kA}) - (X_{IA} - x_{IA})$$

and

$$E = W_a e_{ia} e_{ia}.$$

Setting $\partial E/\partial r_P = 0_P$ gives

$$\begin{aligned} W_a \varepsilon_{iPk} \varepsilon_{ilm} r_l (X_{ka} + x_{ka})(X_{ma} + x_{ma}) \\ = W_a \varepsilon_{iPk} (X_{ka} + x_{ka})(X_{ia} - x_{ia}) \end{aligned}$$

which reduces to

$$2\mathbf{V} = -(\boldsymbol{Q} + \boldsymbol{Q}_0)\mathbf{r}$$

in which

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{M} + \boldsymbol{M}^T - 2\boldsymbol{I}\,\mathrm{tr}\,\boldsymbol{M} \\ \boldsymbol{Q}_0 &= \boldsymbol{S} + \boldsymbol{S}' - \boldsymbol{I}(\mathrm{tr}\,\boldsymbol{S} + \mathrm{tr}\,\boldsymbol{S}') \\ V_I &= \varepsilon_{Ijk} M_{jk} \\ S_{IJ} &= W_a x_{Ia} x_{Ja} \\ S'_{IJ} &= W_a X_{Ia} X_{Ja}. \end{aligned}$$

Thus a direct solution for $\mathbf{r}$ is obtained,

$$\mathbf{r} = -2(\boldsymbol{Q}_0 + \boldsymbol{Q})^{-1}\mathbf{V},$$

the elements of which are u, v and w, and may be used to construct the orthogonal matrix as in Section 3.3.1.2.1. $\boldsymbol{Q} + \boldsymbol{Q}_0$ may be obtained directly from $\boldsymbol{X} + \boldsymbol{x}$.

If the requisite rotation is 180°, $(\boldsymbol{Q}_0 + \boldsymbol{Q})$ is singular and cannot be inverted. In this case any row or column of the adjoint of $(\boldsymbol{Q}_0 + \boldsymbol{Q})$ is a vector in the direction of the axis. Normalizing this vector to unity, giving $\boldsymbol{l}$, gives the requisite orthogonal matrix as

$$\boldsymbol{R} = 2\boldsymbol{ll}^T - \boldsymbol{I}.$$

Note that MacKay's residual E is quadratic in $\mathbf{r}$. E therefore has one minimum and no maximum, and the minimum is reached on the first cycle of least squares. By contrast, the objective function E that is minimized in methods (i), (ii), (v) and (vii) has one minimum, one maximum and two saddle points in the space of the vector $\mathbf{r}$, as shown in (vii).

It may be shown (Diamond, 1989) that if MacKay's solution vector $\mathbf{r}$ is denoted by $\mathbf{r}_\mathrm{M}$ and that given by the other methods [except (iv)] by $\mathbf{r}_\mathrm{O}$ then

$$\mathbf{r}_\mathrm{M} = \mathbf{r}_\mathrm{O} - \boldsymbol{A}^{-1}\boldsymbol{B}\mathbf{r}_\mathrm{O}$$

in which $\boldsymbol{A}$ and $\boldsymbol{B}$ are real symmetric, positive semi-definite. $\boldsymbol{A}$ is positive definite unless all the individual vector sums $(\mathbf{X} + \mathbf{x})$ are parallel, as can happen when the best rotation is 180°. Thus the MacKay method only gives the same result as the other methods if:

(*a*) the initial orientation is optimal, for then $\mathbf{r}_\mathrm{O} = \mathbf{0}$, or

(*b*) perfect fitting is possible, for then $\boldsymbol{B} = \mathbf{0}$, or

(*c*) all the residual vectors (after fitting by $\mathbf{r}_\mathrm{O}$) are parallel to $\mathbf{r}_\mathrm{O}$, for then $\boldsymbol{B}$ is singular such that $\boldsymbol{B}\mathbf{r}_\mathrm{O} = \mathbf{0}$. In general, $|\mathbf{r}_\mathrm{M}| \leq |\mathbf{r}_\mathrm{O}|$. $\mathbf{r}_\mathrm{O}$ may be found by iterating $\mathbf{r}_\mathrm{M}$, but $\boldsymbol{x}$ must be replaced by $\boldsymbol{Rx}$ on each iteration.

(vii) Diamond's second method. This is closely related to MacKay's method, but uses a four-dimensional vector $\boldsymbol{\rho}$ with components λ, μ, ν and σ in which λ, μ and ν are the direction cosines of the rotation axis multiplied by $\sin(\theta/2)$ and σ is $\cos(\theta/2)$. In terms of such a vector Diamond (1988) showed that

$$E = E_0 - 2\boldsymbol{\rho}^T\boldsymbol{P}\boldsymbol{\rho}$$

in which E is the weighted sum of squares of coordinate differences, as before, E_0 is its value before any rotation is applied and $\boldsymbol{P}$ is the matrix

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{Q} & \mathbf{V} \\ \mathbf{V}^T & 0 \end{pmatrix}.$$

The rotation matrix $\boldsymbol{R}$ corresponding to the vector $\boldsymbol{\rho}$ is then the last of the forms for $\boldsymbol{R}$ given in Section 3.3.1.2.1.

The minimum E is therefore E_0 minus twice the largest eigenvalue of $\boldsymbol{P}$ since $\boldsymbol{\rho}^T\boldsymbol{\rho} = 1$, and a stationary value of E occurs when $\boldsymbol{\rho}$ is any of the four eigenvectors of $\boldsymbol{P}$. E thus has a maximum, a minimum and two saddle points, in general, and its value may be determined before any coordinates are transformed. Diamond also showed that the orientations giving these stationary values are related by the operations of 222 symmetry. Equivalent results have also been obtained by Kearsley (1989).

As an alternative to solving a 4×4 eigenproblem, Diamond also showed that the vector $\mathbf{r}$, as in MacKay's solution, may be obtained by iterating

$$\begin{aligned} \alpha_0 &= E_0/2 \\ \mathbf{r}_n &= (\alpha_n \boldsymbol{I} - \boldsymbol{Q})^{-1}\mathbf{V} \\ \alpha_{n+1} &= \frac{\mathbf{V} \cdot \mathbf{r}_n + \alpha_n r_n^2}{1 + r_n^2} \end{aligned}$$

which has the property that if $\boldsymbol{X}$ and $\boldsymbol{x}$ are exactly superposable then α_0 is the exact solution and no iteration is necessary. If $\boldsymbol{X}$ and $\boldsymbol{x}$ are similar but not exactly superposable then a small number of iterations may be required to reach a stable $\mathbf{r}$ vector, though the matrix $\boldsymbol{Q}_0$ is not required. As in MacKay's solution, $(\alpha\boldsymbol{I} - \boldsymbol{Q})$ is singular at the end of the iteration if the required rotation is 180°, but the MacKay and Diamond methods both have the advantage that improper rotations are never generated by these means, and methods based on $\boldsymbol{P}$ and $\boldsymbol{\rho}$ rather than $\boldsymbol{Q}$ and $\mathbf{r}$ are trouble-free

for 180° rotations. The iterative loop in this method does not require $\boldsymbol{Rx}$ to be redetermined on each cycle.

Finally, it may be shown that if p_1, p_2, p_3, p_4 are the eigenvalues of $\boldsymbol{P}$ arranged in descending order and

$$p_1 - p_2 - p_3 + p_4$$

is negative, then a closer superposition may be obtained by reversing the chirality of one of the vector sets, and the $\boldsymbol{R}$ matrix constructed from $\boldsymbol{\rho}_4$ optimally superimposes $\boldsymbol{Rx}$ onto $-\boldsymbol{X}$, the enantiomer of $\boldsymbol{X}$ (Diamond, 1990*b*).

These ideas lend themselves to the problem of the multiple simultaneous superposition of many structures, which is a nonlinear problem with the possibility of multiple solutions [Diamond (1992); see also Kearsley (1990) and Shapiro *et al.* (1992)], and they can provide for cluster analysis based on structural similarity to find subsets of similar structures within an ensemble of structures, such as may arise from NMR calculations (Diamond, 1995).

3.3.1.2.3. Orthogonalization of impure rotations

There are several ways of deriving a strictly orthogonal matrix from a given approximately orthogonal matrix, among them the following.

(i) The Gram–Schmidt process. This is probably the simplest and the easiest to compute. If the given matrix consists of three column vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ (later referred to as primers) which are to be replaced by three column vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ then the process is

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1/|\mathbf{v}_1| \\
\mathbf{u}_2 &= \mathbf{v}_2 - (\mathbf{u}_1 \cdot \mathbf{v}_2)\mathbf{u}_1 \\
\mathbf{u}_2 &= \mathbf{u}_2/|\mathbf{u}_2| \\
\mathbf{u}_3 &= \mathbf{v}_3 - (\mathbf{u}_1 \cdot \mathbf{v}_3)\mathbf{u}_1 - (\mathbf{u}_2 \cdot \mathbf{v}_3)\mathbf{u}_2 \\
\mathbf{u}_3 &= \mathbf{u}_3/|\mathbf{u}_3|.
\end{aligned}$$

As successive vectors are established, each vector $\mathbf{v}$ has subtracted from it its components in the directions of established vectors, and the remainder is normalized. The method will fail at the normalization step if the vectors $\mathbf{v}$ are not linearly independent. Otherwise, the process may be extended to any number of dimensions.

The weakness of the method is that, though $\mathbf{u}_1$ differs from $\mathbf{v}_1$ only in scale, $\mathbf{u}_N$ may differ grossly from $\mathbf{v}_N$ as the various columns are not treated equivalently.

(ii) A preferable method which treats all vectors equivalently is to iteratively replace the matrix $\boldsymbol{M}$ by $\frac{1}{2}(\boldsymbol{M} + \boldsymbol{M}^{T-1})$.

Defining the residual matrix $\boldsymbol{E}$ as

$$\boldsymbol{E} = \boldsymbol{M}\boldsymbol{M}^T - \boldsymbol{I},$$

then on each iteration $\boldsymbol{E}$ is replaced by

$$\boldsymbol{E}^2(\boldsymbol{M}\boldsymbol{M}^T)^{-1}/4$$

and convergence necessarily ensues.

(iii) A third method resolves $\boldsymbol{M}$ into its symmetric and antisymmetric parts

$$\boldsymbol{S} = \tfrac{1}{2}(\boldsymbol{M} + \boldsymbol{M}^T), \quad \boldsymbol{A} = \tfrac{1}{2}(\boldsymbol{M} - \boldsymbol{M}^T), \quad \boldsymbol{M} = \boldsymbol{S} + \boldsymbol{A}$$

and constructs an orthogonal matrix for which only $\boldsymbol{S}$ is altered. $\boldsymbol{A}$ determines l, m, n and θ as shown in Section 3.3.1.2.1, and from these a new $\boldsymbol{S}$ may be constructed.

(iv) A fourth method is to treat the general matrix $\boldsymbol{M}$ as a combination of pure strain and pure rotation. Setting

$$\boldsymbol{M} = \boldsymbol{R}\boldsymbol{T}$$

with $\boldsymbol{R}$ orthogonal and $\boldsymbol{T}$ symmetrical gives

$$\boldsymbol{T} = (\boldsymbol{M}^T\boldsymbol{M})^{1/2}, \quad \boldsymbol{R} = \boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^{-1/2}.$$

The rotation so found is the one which exactly superposes those three mutually perpendicular directions which remain mutually perpendicular under the transformation $\boldsymbol{M}$.

$\boldsymbol{T} - \boldsymbol{I}$ is then the strain tensor of an unrotated body.

Writing $\boldsymbol{M} = \boldsymbol{T}\boldsymbol{R}$, $\boldsymbol{T} = (\boldsymbol{M}\boldsymbol{M}^T)^{1/2}$, $\boldsymbol{R} = (\boldsymbol{M}\boldsymbol{M}^T)^{-1/2}\boldsymbol{M}$ may also be useful, in which $\boldsymbol{T} - \boldsymbol{I}$ is the strain tensor of a rotated body. See also Section 3.3.1.2.2 (iv).

3.3.1.2.4. Eigenvalues and eigenvectors of orthogonal matrices

If $\boldsymbol{R}$ is the orthogonal matrix given in Section 3.3.1.2.1 in terms of the direction cosines l, m and n of the axis of rotation, then it is clear that (l, m, n) is an eigenvector of $\boldsymbol{R}$ with eigenvalue unity because

$$\boldsymbol{R}\begin{pmatrix} l \\ m \\ n \end{pmatrix} = \begin{pmatrix} l \\ m \\ n \end{pmatrix}.$$

Consideration of the determinant $|\boldsymbol{R} - \lambda\boldsymbol{I}| = 0$ shows that the sum of the three eigenvalues is $1 + 2\cos\theta$ and that their product is unity. Hence the three eigenvalues are 1, $e^{i\theta}$ and $e^{-i\theta}$. Since $\boldsymbol{R}$ is real, its product with any real vector is also real, yet its product with an eigenvector must, in general, be complex. Thus the eigenvectors must themselves be complex.

The remaining two eigenvectors $\mathbf{u}$ may be found using the results of Section 3.3.1.2.1 (*q.v.*) according to

$$\boldsymbol{R}\mathbf{u} = \mathbf{u} + \frac{2}{1+t^2}\{(\mathbf{r}\times\mathbf{u}) + [\mathbf{r}\times(\mathbf{r}\times\mathbf{u})]\} = \mathbf{u}e^{\pm i\theta} = \mathbf{u}\frac{1 \pm it}{1 \mp it},$$

which is solved by any vector of the form

$$\mathbf{u} = \mathbf{l}\times\mathbf{v} \mp i\mathbf{l}\times(\mathbf{l}\times\mathbf{v})$$

for any real vector $\mathbf{v}$, where $\mathbf{l}$ is the normalized axis vector, $\mathbf{l}t = \mathbf{r}$, $|\mathbf{l}| = 1$, $t = \tan(\theta/2)$. Eigenvectors for the two eigenvalues may have unrelated $\mathbf{v}$ vectors though the sign choices are coupled. If the vector $\mathbf{v}$ is rotated about $\mathbf{l}$ through an angle φ the corresponding vector $\mathbf{u}$ is multiplied by $e^{-i\varphi}$ and remains an eigenvector. Using superscript signs to denote the sign of θ in the eigenvalue with which each vector is associated, the matrix

$$\boldsymbol{U} = (\mathbf{l},\ \mathbf{u}^+,\ \mathbf{u}^-)$$

has the properties that

$$\boldsymbol{R}\boldsymbol{U} = \boldsymbol{U}\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\theta} & 0 \\ 0 & 0 & e^{-i\theta} \end{pmatrix}$$

and

$$\boldsymbol{U}^{*T}\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2|\mathbf{l}\times\mathbf{v}^+|^2 & 0 \\ 0 & 0 & 2|\mathbf{l}\times\mathbf{v}^-|^2 \end{pmatrix}$$

which places restrictions on $\mathbf{v}$ if this is to be the identity. Note that the 23 element vanishes even in the absence of any relationship between $\mathbf{v}^+$ and $\mathbf{v}^-$.

A convenient form for $\boldsymbol{U}$, symmetrical in the elements of $\mathbf{l}$, is obtained by setting $\mathbf{v}^+ = \mathbf{v}^- = [111]$ and is

$$\boldsymbol{U} = \begin{pmatrix} l & \{(m-n) - i[l(l+m+n)-1]\}/d & \{(m-n) + i[l(l+m+n)-1]\}/d \\ m & \{(n-l) - i[m(l+m+n)-1]\}/d & \{(n-l) + i[m(l+m+n)-1]\}/d \\ n & \{(l-m) - i[n(l+m+n)-1]\}/d & \{(l-m) + i[n(l+m+n)-1]\}/d \end{pmatrix}$$

in which the normalizing denominator is given by

$$d = 2\sqrt{1 - lm - mn - nl}.$$

3.3.1.3. Projection transformations and spaces

In the following section we address the question of the relationship between the coordinates of a molecular model and the corresponding coordinates on the screen of the graphics device. A good introduction to this topic is given by Newman & Sproull (1973), and Foley *et al.* (1990) give a comprehensive account of the field, including recent developments, especially those arising from the development of raster-graphics technologies.

3.3.1.3.1. Definitions

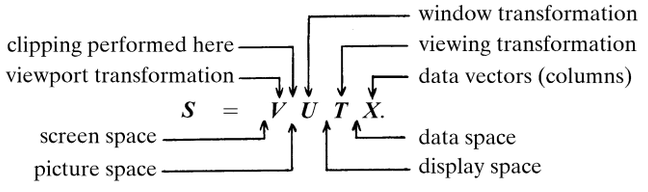
Typically, the coordinates, $\mathbf{X}$, of points in an object to be drawn are held in homogeneous Cartesian form as described in Section 3.3.1.1.2. Such coordinates are said to be in *data space* or world coordinates and this coordinate system is generally a constant aspect of the problem.

In order to view these data in convenient ways such coordinates may be subjected to a 4×4 *viewing transformation* $\boldsymbol{T}$, affecting orientation, scale *etc.*, the resulting coordinates $\boldsymbol{TX}$ being then in *display space*. Here, and throughout what follows, we treat position vectors as columns with transformation matrices as factors on the left, though some writers do the reverse.

In general, only some portion of display space which lies inside a certain frustum of a pyramid is required to fall within the picture. The pyramid may be thought of as having the observer's eye at its vertex, with a rectangular base corresponding to the picture area. This volume is called a *window*. A transformation, $\boldsymbol{U}$, which takes *display-space* coordinates as input and generates vectors (X, Y, Z, W) for which X/W and $Y/W = \pm 1$ for points on the left, right, top and bottom boundaries of the window and for which Z/W takes particular values on the front and back planes of the window, is said to be a *windowing transformation*. In machines for which Z/W controls intensity depth cueing, the range of Z/W corresponding to the window is likely to be 0 to 1 rather than -1 to 1. Coordinates obtained by multiplying display-space coordinates by $\boldsymbol{U}$ are termed *picture-space* coordinates. Mathematically, $\boldsymbol{U}$ is a 4×4 matrix like any other, but functionally it is special. Declaring a transformation to be a windowing transformation implies that only resulting points having $|X|, |Y| < W$ and positive $Z < W$ are to be plotted. Machines with clipping hardware to truncate lines which run out of the picture perform clipping on the output from the windowing transformation.

Finally, the picture has to be drawn in some rectangular portion of the screen which is allocated for the purpose. Such an area is termed a *viewport* and is defined in terms of *screen coordinates* which are defined absolutely for the hardware in question as $\pm n$ for full-screen deflection, where n is declared by the manufacturer. *Screen coordinates* are obtained from picture coordinates with a *viewport transformation*, $\boldsymbol{V}$.[1]

To summarize, screen coordinates, $\mathbf{S}$, are given by



[1] In recent years it has become increasingly common, especially in two-dimensional work, to apply the term 'window' to what is here called a viewport, but in this chapter we use these terms in the manner described in the text.

3.3.1.3.2. Translation

The transformation

$$\begin{pmatrix} N\boldsymbol{I} & \mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix}\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix} = \begin{pmatrix} \mathbf{X}N + \mathbf{V}W \\ NW \end{pmatrix} \simeq \begin{pmatrix} \mathbf{X} + \mathbf{V}W/N \\ W \end{pmatrix} \simeq \begin{pmatrix} \mathbf{X}/W + \mathbf{V}/N \\ 1 \end{pmatrix}$$

evidently corresponds to the addition of the vector $\mathbf{V}W/N$ to the components of $\mathbf{X}$ or of $\mathbf{V}/N$ to the components of $\mathbf{X}/W$. ($\boldsymbol{I}$ is the identity.) Displacements may thus be affected by expressing the required displacement vector in homogeneous coordinates with any suitable choice of N (commonly, $N = W$), with $\mathbf{V}$ scaled to correspond to this choice, and loading the 4×4 transformation matrix as indicated above.

3.3.1.3.3. Rotation

Rotation about the origin is achieved by

$$\begin{pmatrix} N\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & N \end{pmatrix}\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix} = \begin{pmatrix} N\boldsymbol{R}\mathbf{X} \\ NW \end{pmatrix} \simeq \begin{pmatrix} \boldsymbol{R}\mathbf{X} \\ W \end{pmatrix},$$

in which $\boldsymbol{R}$ is an orthogonal 3×3 matrix. $\boldsymbol{R}$ necessarily has elements not exceeding one in modulus. For machines using integer arithmetic, therefore, N would be chosen large enough (usually half the largest possible integer) for the product $N\boldsymbol{R}$ to be well represented in the available word length. Characteristically, N affects resolution but not scale.

3.3.1.3.4. Scale

The transformation

$$\begin{pmatrix} SN\boldsymbol{I} & \mathbf{0} \\ \mathbf{0}^T & N \end{pmatrix}\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix} = \begin{pmatrix} SN\mathbf{X} \\ NW \end{pmatrix} \simeq \begin{pmatrix} S\mathbf{X} \\ W \end{pmatrix}$$

scales the vector $(\mathbf{X}, W)$ by the factor S. For integer working and $|S| < 1$, N should be set to the largest representable integer. For $|S| > 1$ the product SN should be the largest representable integer, N being reduced accordingly.

3.3.1.3.5. Windowing and perspective

It is necessary at this point to relate the discussion to the axial system inherent in the graphics device employed. One common system adopts X horizontal and to the right when viewing the screen, Y vertically upwards in the plane of the screen, and Z normal to X and Y with $+Z$ into the screen. This is, unfortunately, a left-handed system in that $(\mathbf{X} \times \mathbf{Y}) \cdot \mathbf{Z}$ is negative. Since it is usual in crystallographic work to use right-handed axial systems it is necessary to incorporate a matrix of the form

$$\begin{pmatrix} W & 0 & 0 & 0 \\ 0 & W & 0 & 0 \\ 0 & 0 & -W & 0 \\ 0 & 0 & 0 & W \end{pmatrix}$$

either as the left-most factor in the matrix $\boldsymbol{T}$ or as the right-most factor in the windowing transformation $\boldsymbol{U}$ (see Section 3.3.1.3.1). The latter choice is to be preferred and is adopted here. The former choice leads to complications if transformations in display space will be required. Display-space coordinates are necessarily referred to this axial system.

Let L, R, T, B, N and F be the left, right, top, bottom, near and far boundaries of the windowed volume ($L < R$, $T > B$, $N < F$), S be the Z coordinate of the screen, and C, D and E be the coordinates of the observer's eye position, all ten of these parameters being referred to the origin of display space as origin, which may be anywhere in relation to the hardware. L, R, T and B are to be evaluated in the screen plane. All ten parameters may be referred

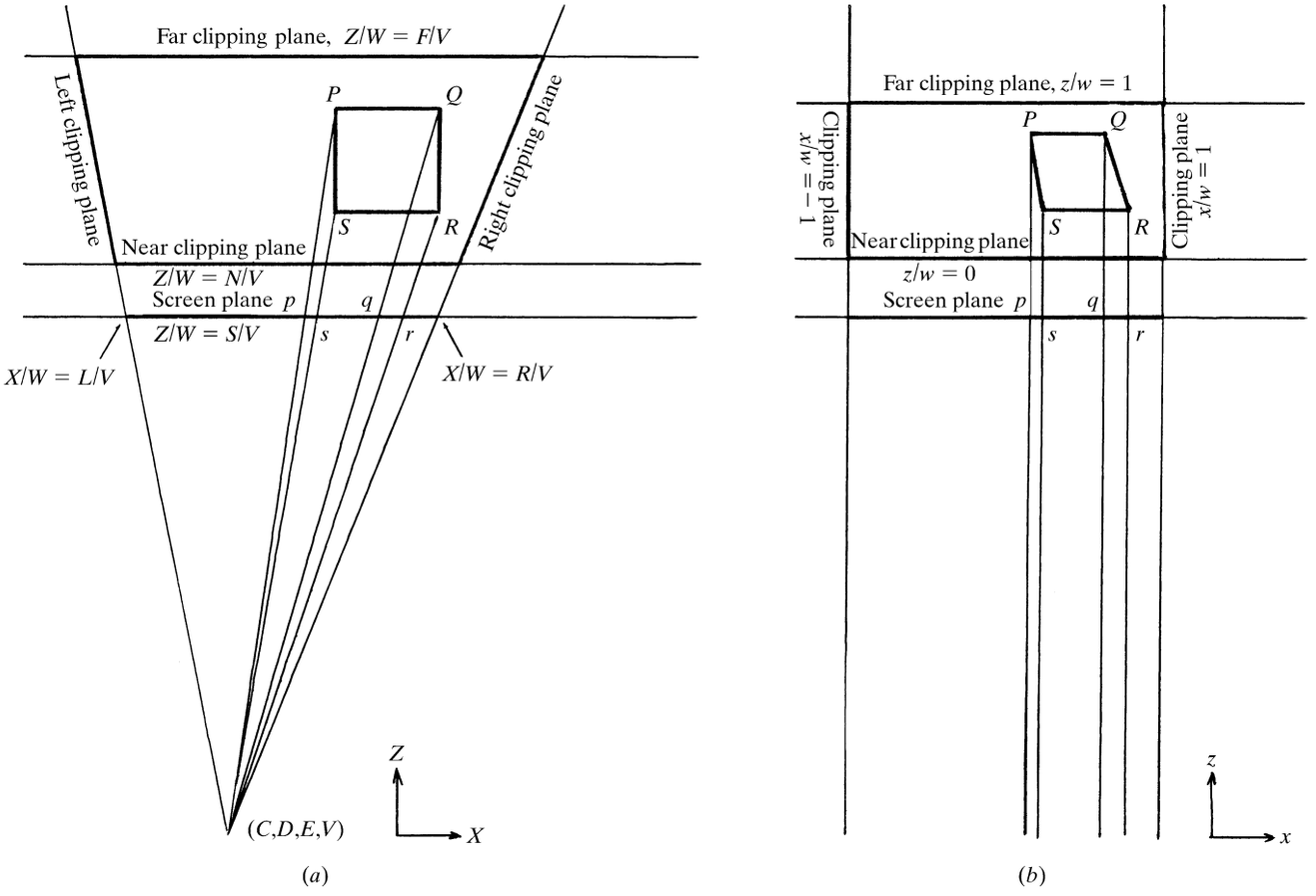


Fig. 3.3.1.1. The relationship between display-space coordinates (X, Y, Z, W) and picture-space coordinates (x, y, z, w) derived from them by the window transformation, $\boldsymbol{U}$. (*a*) Display space (in X, Z projection) showing a square object P, Q, R, S for display viewed from the position (C, D, E, V). The bold trapezium is the window (volume) and the bold line is the viewport portion of the screen. The points P, Q, R and S must be plotted at p, q, r and s to give the correct impression of the object. (*b*) Picture space (in x, z projection). The window is mapped to a rectangle and all sight lines are parallel to the z axis, but the object P, Q, R, S is no longer square. The distribution of p, q, r and s is identical in the two cases. Note that z/w values are not linear on Z/W, and that the origin of picture space arises at the midpoint of the near clipping plane, regardless of the location of the origin of display space. The figure is accurately to scale for coincident viewport positions. The words 'Left clipping plane', if part of the scene in display space, would currently be obscured, but would come into view if the eye moved to the right, increasing C, as the left clipping plane would pivot about the point L/V in the screen plane.

to their own fourth coordinate, V, meaning that the point (X, Y, Z, W) in display space will be on the left boundary of the picture if $X/W = L/V$ when $Z/W = S/V$. V may be freely chosen so that all eleven quantities and all elements of $\boldsymbol{U}$ suit the word length of the machine. These relationships are illustrated in Fig. 3.3.1.1.

Since

$$(X, Y, Z, W) \simeq \left(\frac{XV}{W}, \frac{YV}{W}, \frac{ZV}{W}, V\right),$$

XV/W is a display-space coordinate on the same scale as the window parameters. This must be plotted on the screen at an X coordinate (on the scale of the window parameters) which is the weighted mean of XV/W and C, the weights being $(S - E)$ and $(ZV/W - S)$, respectively. This provides perspective because the weighted mean is at the point where the straight line from (X, Y, Z, W) to the eye intersects the screen. This then has to be mapped into the L-to-R interval, so that picture-space coordinates (x, y, z, w) are given by

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} \frac{2(S-E)V}{(R-L)} & 0 & \frac{(2C-R-L)V}{(R-L)} & \frac{(R+L)E-2SC}{(R-L)} \\ 0 & \frac{2(S-E)V}{(T-B)} & \frac{(2D-T-B)V}{(T-B)} & \frac{(T+B)E-2SD}{(T-B)} \\ 0 & 0 & \frac{(F-E)V}{(F-N)} & \frac{-N(F-E)}{(F-N)} \\ 0 & 0 & V & -E \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}$$

which provides for $|x/w|$ and $|y/w|$ to be unity on the picture boundaries, which is usually a requirement of the clipping hardware, and for $0 < z/w < 1$, zero being for the near-plane boundary. Even though z/w is not linear on Z/W, straight lines and planes in display space transform to straight lines and planes in picture space, the nonlinearity affecting only distances. Thus vector-drawing machines are not disadvantaged by the introduction of perspective.

Note that the dimensionality of X/W must equal that of S/V and that this may be regarded as length or as a pure number, but that in either case x/w is dimensionless, consistent with the stipulation that the picture boundaries be defined by the pure number ± 1.

The above matrix is $\boldsymbol{U}$ and is suited to left-handed hardware systems. Note that only the last column of $\boldsymbol{U}$ (the translational part) is sensitive to the location of the origin of display space and that if the eye is on the normal to the picture centre then $C = \frac{1}{2}(R + L)$, $D = \frac{1}{2}(T + B)$ and simplifications result. If C, D and E can be continuously monitored then dynamic parallax as well as perspective may be obtained (Diamond *et al.*, 1982).

If data space is referred to right-handed axes, the viewing transformation $\boldsymbol{T}$ involves only proper rotations and the hardware uses a left-handed axial system then elements in the third column of $\boldsymbol{U}$ should be negated, as explained in the opening paragraph.

To provide for orthographic projection, multiply every element of $\boldsymbol{U}$ by $-K/E$ and then let $E \rightarrow -\infty$, choosing some positive K

to suit the word length of the machine [see Section 3.3.1.1.2 (iii)]. The result is

$$\boldsymbol{U}' \simeq \begin{pmatrix} \dfrac{2KV}{(R-L)} & 0 & 0 & \dfrac{-K(R+L)}{(R-L)} \\ 0 & \dfrac{2KV}{(T-B)} & 0 & \dfrac{-K(T+B)}{(T-B)} \\ 0 & 0 & \dfrac{KV}{(F-N)} & \dfrac{-KN}{(F-N)} \\ 0 & 0 & 0 & K \end{pmatrix},$$

which is the orthographic window.

It may be convenient in some applications to separate the functions of windowing and the application of perspective, and to write

$$\boldsymbol{U} = \boldsymbol{U}'\boldsymbol{P},$$

where $\boldsymbol{U}$ and $\boldsymbol{U}'$ are as above and $\boldsymbol{P}$ is a perspective transformation given by

$$\boldsymbol{P} = (\boldsymbol{U}')^{-1}\boldsymbol{U} \simeq \begin{pmatrix} S-E & 0 & C & -SC/V \\ 0 & S-E & D & -SD/V \\ 0 & 0 & F-E+N & -NF/V \\ 0 & 0 & V & -E \end{pmatrix},$$

which involves F and N but not R, L, T or B. In this form the action of $\boldsymbol{P}$ may be thought of as compressing distant parts of display space prior to an orthographic projection by $\boldsymbol{U}'$ into picture space.

Other factorizations of $\boldsymbol{U}$ are possible, for example

$$\boldsymbol{U} = \boldsymbol{U}''\boldsymbol{P}'$$

with

$$\boldsymbol{U}'' \simeq \begin{pmatrix} \dfrac{2KV}{R-L} & 0 & 0 & \dfrac{-K(R+L)}{(R-L)} \\ 0 & \dfrac{2KV}{T-B} & 0 & \dfrac{-K(T+B)}{(T-B)} \\ 0 & 0 & \dfrac{KV(N-E)(F-E)}{E^2(F-N)} & \dfrac{KN(F-E)}{E(F-N)} \\ 0 & 0 & 0 & K \end{pmatrix}$$

$$\boldsymbol{P}' \simeq \begin{pmatrix} S-E & 0 & C & -SC/V \\ 0 & S-E & D & -SD/V \\ 0 & 0 & -E & 0 \\ 0 & 0 & V & -E \end{pmatrix},$$

which renders $\boldsymbol{P}'$ independent of all six boundary planes, but $\boldsymbol{U}''$ is no longer independent of E. It is not possible to factorize $\boldsymbol{U}$ so that the left factor is a function only of the boundary planes and the right factor a function only of eye and screen positions.

Note that as $E \to -\infty$, $\boldsymbol{U}'' \to \boldsymbol{U}'$, $\boldsymbol{P}$ and $\boldsymbol{P}' \to -\boldsymbol{I}E \simeq \boldsymbol{I}$.

3.3.1.3.6. Stereoviews

Assuming that left- and right-eye views are to be presented through the same viewport (next section) or that their viewports are to be superimposed by an external optical system, *e.g.* Ortony mirrors, then stereopairs are obtained by using appropriate eye coordinates in the $\boldsymbol{U}$ matrix of the previous section. However, $\boldsymbol{U}$ may be factorized according to

$$\boldsymbol{U} = \boldsymbol{U}'''\boldsymbol{S}$$

in which $\boldsymbol{U}'''$ is the matrix $\boldsymbol{U}$ obtained by setting (C, D, E, V) to correspond to the point midway between the viewer's eyes and

$$\boldsymbol{S} = \begin{pmatrix} 1 & 0 & c/(S-E) & -cS/(S-E)V \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\simeq \begin{pmatrix} V & 0 & cV/(S-E) & -cS/(S-E) \\ 0 & V & 0 & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & V \end{pmatrix}$$

in which $(c, 0, 0, V)$ is the position of the right eye relative to the mean eye position, and the left-eye view is obtained by negating c.

Stereo is often approximated by introducing a rotation about the Y axis of $\pm\sin^{-1}[c/(S-E)]$ to the views or $\sin^{-1}[2c/(S-E)]$ to one of them. The first corresponds to

$$\boldsymbol{S} = \begin{pmatrix} \sqrt{1-\sigma^2} & 0 & \sigma & 0 \\ 0 & 1 & 0 & 0 \\ -\sigma & 0 & \sqrt{1-\sigma^2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with $\sigma = c/(S-E)$. The main difference is in the resulting Z value, which only affects depth cueing and z clipping. The X translation which arises if $S \neq 0$ is also suppressed, but this is not likely to be noticeable. σ is often treated as a constant, such as $\sin 3^\circ$.

The distinction in principle between the true $\boldsymbol{S}$ and the rotational approximation is that with the true $\boldsymbol{S}$ the eye moves relative to the screen and the displayed object, whereas with the approximation the eye and the screen are moved relative to the displayed object, in going from one view to the other.

Strobing of left and right images may conveniently be accomplished with an electro-optic liquid-crystal shutter as described by Harris *et al.* (1985). The shutter is switched by the display itself, thus solving the synchronization problem in a manner free of inertia.

A further discussion of stereopairs may be found in Johnson (1970) and in Thomas (1993), the second of which generalizes the treatment to allow for the possible presence of an optical system.

3.3.1.3.7. Viewports

The window transformation of the previous two sections has been constructed to yield picture coordinates (X, Y, Z, W) (formerly called x, y, z, w) such that a point having X/W or $Y/W = \pm 1$ is on the boundary of the picture, and the clipping hardware operates on this basis. However, the edges of the picture need not be at the edges of the screen and a viewport transformation, $\boldsymbol{V}$, is therefore needed to position the picture in the requisite part of the screen.

$$\boldsymbol{V} = \begin{pmatrix} (r-l)/2 & 0 & 0 & (r+l)/2 \\ 0 & (t-b)/2 & 0 & (t+b)/2 \\ 0 & 0 & n & 0 \\ 0 & 0 & 0 & n \end{pmatrix},$$

where r, l, t and b are now the right, left, top and bottom boundaries of the picture area, or viewport, expressed in screen coordinates, and n is the full-screen deflection value. Thus a point with $X/W = 1$ in picture space plots on the screen with an X coordinate which is a fraction r/n of full-screen deflection to the

right. Z/W is unchanged by $\boldsymbol{V}$ and is used only to control intensity in a technique known as depth cueing.

It is necessary, of course, to arrange for the aspect ratio of the viewport, $(r-l)/(t-b)$, to equal that of the window otherwise distortions are introduced.

3.3.1.3.8. Compound transformations

In this section we consider the viewing transformation $\boldsymbol{T}$ of Section 3.3.1.3.1 and its construction in terms of translation, rotation and scaling, Sections 3.3.1.3.2–4. We use $\boldsymbol{T}'$ to denote a new transformation in terms of the prevailing transformation $\boldsymbol{T}$.

We note first that any 4×4 matrix of the form

$$\begin{pmatrix} U\boldsymbol{R} & \mathbf{V} \\ \mathbf{0}^T & W \end{pmatrix},$$

with U a scalar, may be factorized according to

$$\begin{pmatrix} U\boldsymbol{R} & \mathbf{V} \\ \mathbf{0}^T & W \end{pmatrix} \simeq \begin{pmatrix} U\boldsymbol{I} & \mathbf{0} \\ \mathbf{0}^T & W \end{pmatrix}\begin{pmatrix} U\boldsymbol{I} & \mathbf{V} \\ \mathbf{0}^T & U \end{pmatrix}\begin{pmatrix} U\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & U \end{pmatrix}$$

and also that multiplying

$$\begin{pmatrix} U\boldsymbol{R} & \mathbf{V} \\ \mathbf{0}^T & W \end{pmatrix}$$

by an isotropic scaling matrix, a rotation, or a translation, either on the left or on the right, yields a product matrix of the same form, and its inverse

$$\begin{pmatrix} W\boldsymbol{R}^T & -\boldsymbol{R}^T\mathbf{V} \\ \mathbf{0}^T & U \end{pmatrix}$$

is also of this form, *i.e.* any combination of these three operations in any order may be reduced by the above factorization to a rotation about the original origin, a translation (which defines a new origin) and an expansion or contraction about the new origin, applied in that order.

If

$$\begin{pmatrix} N\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & N \end{pmatrix}$$

is a rotation matrix as in Section 3.3.1.3.3, its application produces a rotation about an axis through the origin defined only in the space in which it is applied. For example, if

$$\boldsymbol{R} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{T}'\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix} = \boldsymbol{T}\begin{pmatrix} N\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & N \end{pmatrix}\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix}$$

rotates the image about the z axis of data space, whatever the prevailing viewing transformation, $\boldsymbol{T}$.

Forming

$$\begin{pmatrix} N\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & N \end{pmatrix}\boldsymbol{T}\begin{pmatrix} \mathbf{X} \\ W \end{pmatrix}$$

rotates the image about the z axis of display space, *i.e.* the normal to the tube face under the usual conventions, whatever the prevailing $\boldsymbol{T}$. Furthermore, if this rotation is to appear to be about some chosen position in the picture, *e.g.* the centre, then the window transformation $\boldsymbol{U}$, Section 3.3.1.3.5, must place the origin of display space there by setting $F > S = R + L = T + B = 0 > N$, in the notation of that section.

If a rotation is to be about a point

$$\begin{pmatrix} \mathbf{V} \\ N \end{pmatrix}$$

then

$$\begin{aligned} \boldsymbol{T}' &= \begin{pmatrix} N\mathbf{I} & \mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix}\begin{pmatrix} N'\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & N' \end{pmatrix}\begin{pmatrix} N\boldsymbol{I} & -\mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix}\boldsymbol{T} \\ &\simeq \begin{pmatrix} N\boldsymbol{R} & \mathbf{V} - \boldsymbol{R}\mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix}\boldsymbol{T} \end{aligned}$$

or

$$\begin{aligned} \boldsymbol{T}' &= \boldsymbol{T}\begin{pmatrix} N\boldsymbol{I} & \mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix}\begin{pmatrix} N'\boldsymbol{R} & \mathbf{0} \\ \mathbf{0}^T & N' \end{pmatrix}\begin{pmatrix} N\boldsymbol{I} & -\mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix} \\ &\simeq \boldsymbol{T}\begin{pmatrix} N\boldsymbol{R} & \mathbf{V} - \boldsymbol{R}\mathbf{V} \\ \mathbf{0}^T & N \end{pmatrix} \end{aligned}$$

according to whether $\boldsymbol{R}$ and $\mathbf{V}$ are both defined in display space or both in data space. If the rotation is defined in display space and the position of the centre of rotation is defined in data space, then the first form of $\boldsymbol{T}'$ must be used, in which $\mathbf{V}$ is first computed from

$$\begin{pmatrix} \mathbf{V} \\ N \end{pmatrix} = \boldsymbol{T}\begin{pmatrix} \mathbf{U} \\ W \end{pmatrix}$$

for a rotation centre at

$$\begin{pmatrix} \mathbf{U} \\ W \end{pmatrix}$$

in data space.

For continuous rotations defined in display space it is usually worthwhile to bring the centre of rotation to the origin of display space and leave it there, *i.e.* to omit the left-most factor in the first expression for $\boldsymbol{T}'$. Incremental rotations can then be made by further rotational factors on the left without further attention to $\mathbf{V}$. When continuous rotations are implemented by repeated multiplication of $\boldsymbol{T}$ by a rotation matrix, say thirty times a second for a minute or so, the orthogonality of the top-left partition of $\boldsymbol{T}$ may become degraded by accumulation of round-off error and this should be corrected occasionally by one of the methods of Section 3.3.1.2.3.

It is sometimes a requirement, depending on hardware capabilities, to effect a transformation in display space when access to data space is all that is readily available. In such a case

$$\boldsymbol{T}' = \boldsymbol{T}_1\boldsymbol{T} = \boldsymbol{T}\boldsymbol{T}_2,$$

where $\boldsymbol{T}_1$ is the required alteration to the prevailing viewing transformation $\boldsymbol{T}$ and $\boldsymbol{T}_2$ is the data-space equivalent,

$$\begin{aligned} \boldsymbol{T}_2 = \boldsymbol{T}^{-1}\boldsymbol{T}_1\boldsymbol{T} &= \begin{pmatrix} U\boldsymbol{R} & \mathbf{V} \\ \mathbf{0}^T & W \end{pmatrix}^{-1}\begin{pmatrix} U_1\boldsymbol{R}_1 & \mathbf{V}_1 \\ \mathbf{0}^T & W_1 \end{pmatrix}\begin{pmatrix} U\boldsymbol{R} & \mathbf{V} \\ \mathbf{0}^T & W \end{pmatrix} \\ &\simeq \begin{pmatrix} UU_1\boldsymbol{R}^T\boldsymbol{R}_1\boldsymbol{R} & \boldsymbol{R}^T(U_1\boldsymbol{R}_1\mathbf{V} + W\mathbf{V}_1 - \boldsymbol{W}_1\mathbf{V}) \\ \mathbf{0}^T & UW_1 \end{pmatrix}. \end{aligned}$$

An important special case is when $\boldsymbol{T}_1$ is to effect a rotation about the origin of display space without change of scale, so that $\mathbf{V}_1 = 0$, $U_1 = W_1 = W$, for then

$$\boldsymbol{T}_2 \simeq \begin{pmatrix} U\boldsymbol{R}^T\boldsymbol{R}_1\boldsymbol{R} & \boldsymbol{R}^T(\boldsymbol{R}_1 - \boldsymbol{I})\mathbf{V} \\ \mathbf{0}^T & U \end{pmatrix}.$$

If $\mathbf{r}$ is the required axis of rotation of $\boldsymbol{R}_1$ in display space then the axis of rotation of $\boldsymbol{R}^T\boldsymbol{R}_1\boldsymbol{R}$ in data space is $\mathbf{s} = \boldsymbol{R}^T\mathbf{r}$ since $\boldsymbol{R}^T\boldsymbol{R}_1\boldsymbol{R}\mathbf{s} = \mathbf{s}$. This gives a particularly simple result if $\boldsymbol{R}_1$ is to be a

primitive rotation for then **s** is the relevant row of $\boldsymbol{R}$, and $\boldsymbol{R}^T\boldsymbol{R}_1\boldsymbol{R}$ can be constructed directly from this and the required angle of rotation.

3.3.1.3.9. Inverse transformations

It is frequently a requirement to be able to identify a feature or position in data space from its position on the screen. Facilities for identifying an existing feature on the screen are in many instances provided by the manufacturer as a 'hit' function which correlates the position indicated on the screen by the user (with a tablet or light pen) with the action of drawing and flags the corresponding item in the drawing internally as having been hit. In other instances it may be necessary to be able to indicate a position in data space independently of any drawn feature and this may be done by setting two or more nonparallel sight lines through the displayed volume and finding their best point of intersection in data space.

In Section 3.3.1.3.1 the relationship between data-space coordinates and screen-space coordinates was given as

$$\mathbf{S} = \boldsymbol{VUTX};$$

hence data-space coordinates are given by

$$\mathbf{X} = \boldsymbol{T}^{-1}\boldsymbol{U}^{-1}\boldsymbol{V}^{-1}\mathbf{S}.$$

A line of sight through the displayed volume passing through the point

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

on the screen is the line joining the two position vectors

$$\boldsymbol{S} = \begin{pmatrix} x & x \\ y & y \\ o & n \\ n & n \end{pmatrix}$$

in screen-space coordinates, as in Section 3.3.1.3.7, from which the corresponding two points in data space may be obtained using

$$\boldsymbol{V}^{-1} \simeq \begin{pmatrix} \dfrac{2n}{r-l} & 0 & 0 & \dfrac{-(r+l)}{(r-l)} \\ 0 & \dfrac{2n}{t-b} & 0 & \dfrac{-(t+b)}{(t-b)} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\boldsymbol{U}^{-1} \simeq \begin{pmatrix} \dfrac{R-L}{2(S-E)} & 0 & \dfrac{-C(F-N)}{(F-E)(N-E)} & \dfrac{(R+L)(N-E)-2C(N-S)}{2(N-E)(S-E)} \\ 0 & \dfrac{T-B}{2(S-E)} & \dfrac{-D(F-N)}{(F-E)(N-E)} & \dfrac{(T+B)(N-E)-2D(N-S)}{2(N-E)(S-E)} \\ 0 & 0 & \dfrac{-E(F-N)}{(F-E)(N-E)} & \dfrac{N}{(N-E)} \\ 0 & 0 & \dfrac{-V(F-N)}{(F-E)(N-E)} & \dfrac{V}{(N-E)} \end{pmatrix}$$

in the notation of Section 3.3.1.3.5, and $\boldsymbol{T}^{-1}$ was given in Section 3.3.1.3.8. If orthographic projection is being used ($E = -\infty$) then $\boldsymbol{U}^{-1}$ simplifies to

$$\boldsymbol{U'}^{-1} \simeq \begin{pmatrix} \dfrac{R-L}{2} & 0 & 0 & \dfrac{R+L}{2} \\ 0 & \dfrac{T-B}{2} & 0 & \dfrac{T+B}{2} \\ 0 & 0 & F-N & N \\ 0 & 0 & 0 & V \end{pmatrix}.$$

Each of these inverse matrices may be suitably scaled to suit the word length of the machine [Section 3.3.1.1.2 (iii)].

Having determined the end points of one sight line in data space the viewing transformation $\boldsymbol{T}$ may then be changed and the required position marked again through the screen in the new orientation. Each such operation generates a pair of points in data space, expressed in homogeneous form, with a variety of values for the fourth coordinate. Each such point must then be converted to three dimensions in the form (X/W, Y/W, Z/W), and for each sight line any (three-dimensional) point $\mathbf{p}_A$ on the line and the direction $\mathbf{q}_A$ of the line are established. For each sight line a rank 2 projector matrix $\boldsymbol{M}_A$ of order 3 is formed as

$$\boldsymbol{M}_A = \boldsymbol{I} - \mathbf{q}_A\mathbf{q}_A^T/\mathbf{q}_A^T\mathbf{q}_A$$

and the best point of intersection of the sight lines is given by

$$\left(\sum_a \boldsymbol{M}_a\right)^{-1}\left(\sum_a \boldsymbol{M}_a\mathbf{p}_a\right),$$

to which three-vector a fourth coordinate of unity may be applied.

3.3.1.3.10. The three-axis joystick

The three-axis joystick is a device which depends on compound transformations for its exploitation. As it is usually mounted it consists of a vertical shaft, mounted at its lower end, which can rotate about its own length (the Y axis of display space, Section 3.3.1.3.1), its angular setting, φ, being measured by a shaft encoder in its mounting. At the top of this shaft is a knee-joint coupling to a second shaft. The first angle φ is set to zero when the second shaft is in some selected direction, *e.g.* normal to the screen and towards the viewer, and goes positive if the second shaft is moved clockwise when seen from above. The knee joint itself contains a shaft encoder, providing an angle, ψ, which may be set to zero when the second shaft is horizontal and goes positive when its free end is raised. A knob on the tip of the second shaft can then rotate about an axis along the second shaft, driving a third shaft encoder providing an angle θ. The device may then be used to produce a rotation of the object on the screen about an axis parallel to the second shaft through an angle given by the knob. The necessary transformation is then

$$\begin{aligned} \boldsymbol{R} = &\begin{pmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix} \\ &\times \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix} \\ &\times \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix} \end{aligned}$$

which is

$$\begin{pmatrix} c^2\psi s^2\varphi + (1 - c^2\psi s^2\varphi)c\theta & -s\psi c\psi s\varphi(1 - c\theta) - c\psi c\varphi s\theta & -c^2\psi s\varphi c\varphi(1 - c\theta) + s\psi s\theta \\ -s\psi c\psi s\varphi(1 - c\theta) + c\psi c\varphi s\theta & s^2\psi + c^2\psi c\theta & s\psi c\psi c\varphi(1 - c\theta) + c\psi s\varphi s\theta \\ -c^2\psi s\varphi c\varphi(1 - c\theta) - s\psi s\theta & s\psi c\psi c\varphi(1 - c\theta) - c\psi s\varphi s\theta & c^2\psi c^2\varphi + (1 - c^2\psi c^2\varphi)c\theta \end{pmatrix}$$

in which cos and sin are abbreviated to c and s, which is the standard form with $l = -\cos\psi\sin\varphi$, $m = \sin\psi$, $n = \cos\psi\cos\varphi$.

3.3.1.3.11. Other useful rotations

If rotations in display space are to be controlled by trackerball or tablet then there are two measures available, an x and a y, which can define an axis of rotation in the plane of the screen and an angle θ. If x and y are suitably scaled coordinates of a pen on a tablet then the rotation

$$\begin{pmatrix} \dfrac{y^2 + x^2c}{x^2 + y^2} & \dfrac{-xy(1 - c)}{x^2 + y^2} & x\sqrt{x^2 + y^2} \\ \dfrac{-xy(1 - c)}{x^2 + y^2} & \dfrac{x^2 + y^2c}{x^2 + y^2} & y\sqrt{x^2 + y^2} \\ -x\sqrt{x^2 + y^2} & -y\sqrt{x^2 + y^2} & c \end{pmatrix}$$

with $c = \sqrt{1 - (x^2 + y^2)^2}$ is about an axis in the xy plane (*i.e.* the screen face) normal to (x, y) and with $\sin\theta = x^2 + y^2$. Applied repetitively this gives a quadratic velocity characteristic. Similarly, if an atom at (x, y, z, w) in display space is to be brought onto the z axis by a rotation with its axis in the xy plane the necessary matrix, in homogeneous form, is

$$\begin{pmatrix} \dfrac{x^2z + y^2r}{x^2 + y^2} & \dfrac{-xy(r - z)}{x^2 + y^2} & -x & 0 \\ \dfrac{-xy(r - z)}{x^2 + y^2} & \dfrac{x^2r + y^2z}{x^2 + y^2} & -y & 0 \\ x & y & z & 0 \\ 0 & 0 & 0 & r \end{pmatrix}$$

with $r = \sqrt{x^2 + y^2 + z^2}$.

3.3.1.3.12. Symmetry

In Section 3.3.1.1.1 it was pointed out that it is usual to express coordinates for graphical purposes in Cartesian coordinates in ångström units or nanometres. Symmetry, however, is best expressed in crystallographic fractional coordinates. If a molecule, with Cartesian coordinates, is being displayed, and a symmetry-related neighbour is also to be displayed, then the data-space coordinates must be multiplied by

$$\begin{pmatrix} \boldsymbol{W} & \mathbf{T} \\ \mathbf{0}^T & W \end{pmatrix}\begin{pmatrix} \boldsymbol{M} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}\mathscr{S}\begin{pmatrix} \boldsymbol{M}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}\begin{pmatrix} \boldsymbol{W} & -\mathbf{T} \\ \mathbf{0}^T & W \end{pmatrix},$$

where

$$\begin{pmatrix} \mathbf{T} \\ W \end{pmatrix}$$

are the data-space coordinates of the crystallographic origin, $\boldsymbol{M}$ and $\boldsymbol{M}^{-1}$ are as in Section 3.3.1.1.1 and $\mathscr{S}$ is a crystallographic symmetry operator in homogeneous coordinates, expressed relative to the same crystallographic origin.

For example, in $P2_1$ with the origin on the screw dyad along **b**,

$$\mathscr{S} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} \boldsymbol{M} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}\mathscr{S}\begin{pmatrix} \boldsymbol{M}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{2}b \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$\mathscr{S}$ comprises a proper or improper rotational partition, $\boldsymbol{S}$, in the upper-left 3×3 in the sense that $\boldsymbol{MSM}^{-1}$ is orthogonal, and with the associated fractional lattice translation in the last column, with the last row always consisting of three zeros and 1 at the 4, 4 position. See *IT* A (2005, Chapters 5.2 and 8.1) for a fuller discussion of symmetry using augmented (*i.e.* 4×4) matrices.

3.3.1.4. Modelling transformations

The two sections under this heading are concerned only with the graphical aspects of conformational changes. Determination of such changes is considered under Section 3.3.2.2.

3.3.1.4.1. Rotation about a bond

It is a common requirement in molecular modelling to be able to rotate part of a molecule relative to the remainder about a bond between two atoms.

If four atoms are bonded 1–2–3–4 then the dihedral angle in the bond 2–3 is zero if the four atoms are *cis* planar, and a rotation in the 2–3 bond is, by convention (IUPAC–IUB Commission on Biochemical Nomenclature, 1970), positive if, when looking along the 2–3 bond, the far end rotates clockwise relative to the near end. This is valid for either viewing direction. This sign convention, when applied to the $\boldsymbol{R}$ matrix of Section 3.3.1.2.1, leads to the following statement.

If one of the two atoms is selected as the near atom and the direction cosines are those of the vector from the near atom to the far atom, and if the matrix is to rotate material attached to the far atom (with the reference axes fixed), then a positive rotation in the foregoing sense is generated by a positive θ.

Rotation about a bond normally involves compounding $\boldsymbol{R}$ with translations in the manner of Section 3.3.1.3.8.

3.3.1.4.2. Stacked transformations

A flexible molecule may require to be drawn in any of a number of conformations which are related to one another by, for example, rotations about single bonds, changes of bond angles or changes of bond lengths, all of which changes may be brought about by the application of suitable homogeneous transformations during the drawing of the molecule (Section 3.3.1.3.8). With suitable organization, this may be done without necessarily altering the coordinates of the atoms in the coordinate list, only the transformations being manipulated during drawing.

The use of transformations in the manner shown below is straightforward for simply connected structures or structures containing only rigid rings. Flexible rings may be similarly handled provided that the matrices employed are consistent with the consequential constraints as described in Section 3.3.2.2.1, though this requirement may make real-time folding of flexible rings difficult.

Any simply connected structure may be organized as a tree with a node at each branch point and with an arbitrary number of sites of conformational change between one node and the next. We shall call such sites and their associated matrices 'conformons'. The technique then depends on the stacking technique in which matrices are stored and later recovered in the reverse order of their storage.

One begins at some reference point deemed to be fixed in data space and at this point one stacks the prevailing viewing transformation. From this reference point one advances through the molecule along the structural tree and as each conformon is encountered its matrix is calculated. The product of the prevailing matrix with the conformon matrix is formed and stacked, and this product becomes the prevailing matrix. This product is constructed with the conformon matrix as a factor on the right, *i.e.* in data space as defined in Section 3.3.1.3.1, and is calculated using the coordinates of the molecule in their unmodified form, *i.e.* before any shape changes are brought about.

This progression leads eventually to an extremity of the tree. At this point drawing is commenced using the prevailing matrix and working backwards towards the fixed root, unstacking (or 'popping') a matrix as each conformon is passed until a node is reached, which, in general, will occur only part way back to the root. On reaching such a node drawing is suspended and one advances along the newly found branch as before, stacking matrices, until another extremity is reached when drawing towards the root is resumed. This alternation of stacking matrices while moving away from the root and drawing and unstacking matrices while moving towards the root is continued until the whole tree is traversed.

This process is illustrated schematically in Fig. 3.3.1.2 for a simple tree with one node, numbered 1, and three conformons at *a*, *b* and *c*. One enters the tree with a current viewing transformation $\boldsymbol{T}$ and progresses upwards from the fixed lower extremity. When the conformon at *a* is encountered, $\boldsymbol{T}$ is stacked and the product $\boldsymbol{TM}_a$ is formed. Continuing up the tree, at node 1 either branch may be chosen; we choose the left and, on reaching *b*, $\boldsymbol{TM}_a$ is stacked and $\boldsymbol{TM}_a\boldsymbol{M}_b$ is formed. On reaching the tip drawing down to *b* is done with this transformation, $\boldsymbol{TM}_a$ is then unstacked and drawing continues with this matrix until node 1 is reached. The other branch is then followed to *c* whereupon $\boldsymbol{TM}_a$ is again stacked and the product $\boldsymbol{TM}_a\boldsymbol{M}_c$ is formed. From the tip down as far as *c* is drawn with this matrix, whereupon $\boldsymbol{TM}_a$ is unstacked and drawing continues down to *a*, where $\boldsymbol{T}$ is unstacked before drawing the section nearest the root.

With this organization the matrices associated with *b* and *c* are unaffected by changes in the conformation at *a*, notwithstanding the fact that changes at *a* alter the direction of the axis of rotation at *b* or *c*.

Two other approaches are also possible. One of these is to start at the tip of the left branch, replace the coordinates of atoms between *b* and the tip by $\boldsymbol{M}_b\mathbf{X}$, and later replace all coordinates between the tip and *a* by $\boldsymbol{M}_a\mathbf{X}$, with a similar treatment for the other branch. The advantage of this is that no storage is required for stacked matrices, but the disadvantage is that atoms near the tips of the tree have to be reprocessed for every conformon. It also modifies the stored coordinates, which may or may not be desirable.

The second alternative is to draw upwards from the root using $\boldsymbol{T}$ until *a* is reached, then using $\boldsymbol{TM}_a$ until *b* is reached, then using $\boldsymbol{TM}'_b\boldsymbol{M}_a$ to the tip, but in this formulation $\boldsymbol{M}'_b$ must be based on the geometry that exists at *b* after the transformation $\boldsymbol{M}_a$ has been applied to this region of the molecule, *i.e.* $\boldsymbol{M}'_b$ is characteristic of the final conformation rather than the initial one.

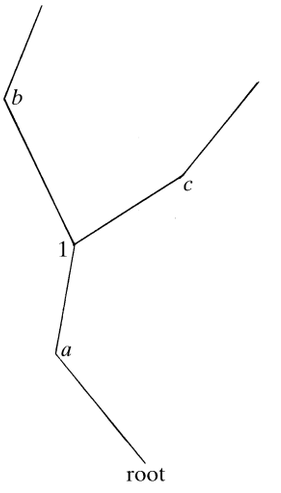


Fig. 3.3.1.2. Schematic representation of a simple branched-chain molecule with a stationary root and two extremities. The positions marked *a*, *b* and *c* are the loci of possible conformational change, here called conformons, and there is a single, numbered branch point.

3.3.1.5. Drawing techniques

3.3.1.5.1. Types of hardware

There are two main types of graphical hardware in use for interactive work, in addition to plotters used for batch work. These main types are raster and vector. In raster equipment the screen is scanned as in television, with a grid of points, called pixels, addressed sequentially as the scan proceeds. Associated with each pixel is a word of memory, usually containing something in the range of 1 to 24 bits per pixel, which controls the colour and intensity to be displayed. Many computer terminals have one bit per pixel (said to be 'single-plane' systems) and these are essentially monochrome and have no grey scale. Four-plane systems are cheap and popular and commonly provide 4-bit by 4-bit look-up tables between the pixel memory and the monitor with one such table for each of the colours red, green and blue. If these tables are each loaded identically then 16 levels of monochrome grey scale are available, but if they are loaded differently 16 different colours are available simultaneously chosen from a total of 4096 possibilities. Four-plane systems are adequate for many applications where colour is used for coding, but are inadequate if colour is intended also to provide realism, where brilliance and saturation must be varied as well as hue. For these applications eight-plane systems are commonly used which permit 256 colours chosen from 16 million using three look-up tables, though the limitations of these can also be felt and full colour is only regarded as being available in 24-plane systems.

Raster-graphics devices are ideal for drawing objects represented by opaque surfaces which can be endowed with realistic reflecting properties (Max, 1984) and they have been successfully used to give effects of transparency. They are also capable of representing shadows, though these are generally difficult to calculate (see Section 3.3.1.5.5). Many devices of this type provide vectorization, area fill and anti-aliasing. Vectorization provides automatically for the loading of relevant pixels on a straight line between specified points. Area fill automatically fills any irregular pre-defined polygon on the screen with a uniform colour with the user specifying only the colour and one point within the polygon. Anti-aliasing is the term used for a technique which softens the staircase effect that may be seen on a line which runs at a small angle to a vertical or horizontal row of pixels.

The main drawback with this type of equipment is that it is slow compared to vector machines. Only relatively simple objects can be displayed with smooth rotation in real time as transformed coordinates have to be converted to pixel addresses and the previous frame needs to be deleted with each new frame unless it is known that each new frame will specify every pixel. However, the technology is advancing rapidly and these restrictions are already disappearing.

Vector machines, on the other hand, are specialized to drawing straight lines between specified points by driving the electron beam along such lines. No time is wasted on blank areas of the screen. Dots may be drawn with arbitrary coordinates, in any order, but areas, if they are to be filled, must be done with a ruling technique which is very seldom done. Images produced by vector

machines are naturally transparent in that foreground does not obscure background, which makes them ideal for seeing into representations of molecular structure.

3.3.1.5.2. Optimization of line drawings

A line drawing consisting of n line segments may be specified by anything from $(n+1)$ to $2n$ position vectors depending on whether the lines are end-to-end connected or independent. Appreciable gains in both processing time and storage requirements may be made in complicated drawings by arranging for line segments to be end-to-end connected as far as possible, and an algorithm for doing this is outlined below. For further details see Diamond (1984*a*).

Supposing that a list of nodal points (atoms if a covalent skeleton is being drawn) exists within a computer with each node appearing only once and that the line segments to be drawn between them are already determined, then at each node there are, generally, both forward and backward connections, forward connections being those to nodes further down the list. A quantity D is calculated at each node which is the number of forward connections minus the number of backward connections. At the commencement of drawing, the first connected node in the list must have a positive D, the last must have a negative D, the sum of all D values must be zero and the sum of the positive ones is the number of strokes required to draw the drawing, a 'stroke' being a sequence of end-to-end connected line segments drawn without interruption. The total number of position vectors required to specify the drawing is then the number of nodes plus the number of strokes plus the number of rings minus one.

Drawing should then be done by scanning the list of nodes from the top looking for a positive D (usually found at the first node), commencing a stroke at this node and decrementing its D value by 1. This stroke is continued from node to node using the specified connections until a negative D is encountered, at which point the stroke is terminated and the D value at the terminating node is incremented by 1. This is done even though this terminating node may also possess some forward connections, as the total number of strokes required is not minimized by keeping a stroke going as far as possible, but by terminating a stroke as soon as it reaches a node at which *some* stroke is bound to terminate.

The next stroke is initiated by resuming the scan for positive D values at the point in the node list where the previous stroke began. If this scan encounters a zero D value at a node which has not hitherto been drawn to, or drawn from, then the node concerned is isolated and not connected to any other, and such nodes may require to be drawn with some special symbol. The expression already given for the number of vectors required is valid in the presence of isolated nodes if drawing an isolated node is allowed one position vector, this vector not being counted as a stroke.

The number of strokes generated by this algorithm is sensitive to the order in which the nodes are listed, but if this resembles a natural order then the number of strokes generated is usually close to the minimum, which is half the number of nodes having an odd number of connections. For example, the letter E has six nodes, four of which have an odd number of connections, so it may be drawn with two strokes.

3.3.1.5.3. Representation of surfaces by lines

The commonest means of representing surfaces, especially contour surfaces, is to consider evenly spaced serial sections and to perform two-dimensional contouring on each section. Repeating this on serial sections in two other orientations then provides a good representation of the surface in three dimensions when all such contours are displayed. The density is normally cited on a grid with submultiples of **a**, **b** and **c** as grid vectors, inverse linear interpolation being used between adjacent grid points to locate points on the contour. For vector-graphics applications it is expedient to connect such points with straight lines; some equipment may be capable of connecting them with splines though this is burdensome or impossible if real-time rotation of the scene is required. Precalculation of splines stored as short vectors is always possible if the proliferation of vectors is acceptable. For efficient drawing it is necessary for the line segments of a contour to be end-to-end connected, which means that it is necessary to contour by following contours wherever they go and not by scanning the grid. Algorithms which function in this way have been given by Heap & Pink (1969) and Diamond (1982*a*). Contouring by grid scanning followed by line connection by the methods of the previous section would be possible but less efficient. Further contouring methods are described by Sutcliffe (1980) and Cockrell (1983).

For raster-graphics devices there is little disadvantage in using curved contours though many raster devices now have vectorizing hardware for loading a line of pixels given only the end points. For these devices well shaped contours may be computed readily, using only linear arithmetic and a grid-scanning approach (Gossling, 1967). Others have colour-coded each pixel according to the density, which provides a contoured visual impression without performing contouring (Hubbard, 1983).

3.3.1.5.4. Representation of surfaces by dots

Connolly (Langridge *et al.*, 1981; Connolly, 1983*a*,*b*) represents surfaces by placing dots on the surface with an approximately uniform superficial density. Connolly's algorithm was developed to display solvent-accessible surfaces of macromolecules and provides for curved concave portions where surface atoms meet. Pearl & Honegger (1983) have developed a similar algorithm, based on a grid, which generates only convex portions which meet in cusps, but is faster to compute than the Connolly surface. Bash *et al.* (1983) have produced a van der Waals surface algorithm fast enough to permit real-time changes to the structure without tearing the surface.

It has become customary to use a dot representation to display computed surfaces, such as the surface at a van der Waals radius from atomic centres, and to use lines to represent experimentally determined surfaces, especially density contours.

3.3.1.5.5. Representation of surfaces by shading

Many techniques have been developed, mainly for raster-graphics devices, for representing molecular surfaces and these have been very well reviewed by Max (1984).

The simplest technique in this class consists in representing each atom by a uniform disc, or high polygon, which can be colour-coded and area-filled by the firmware of the device. If such atoms are sorted on their z coordinate and drawn in order, furthest ones first, so that nearer ones partly or completely overwrite the further ones then the result is a simple representation of the molecule as seen from the front. This technique is fast and has its uses when a rapid schematic is all that is required. In one sense it is wasteful to process distant atoms when they are going to be overwritten by foreground atoms, but front-to-back processing requires the boundaries of visible parts of partially obscured atoms near the front to be determined before they can be painted or, alternatively, every pixel must be tested before loading to see if it is already loaded. Not only does this approach give a uniform rendering over the whole area of one atom, it also gives a boundary between overlapping atoms with almost equal z values which completes the circle of the nearer atom, though it should be an arc of an ellipse when the atoms are drawn with radii exceeding their covalent radii.

Greater realism is achieved by establishing a z buffer, which is an additional area of memory with one word per pixel, in which is stored the z value of the currently loaded feature in each pixel.

Treatments which take account of the sphericity are then possible and correct arcs of intersection for interpenetrating spheres and more complicated entities arise naturally through loading a colour value into a pixel only if the z coordinate is less than that of the currently loaded value. This z buffer and the associated x, y coordinates should be in picture space or screen space rather than display space since only after the application of perspective can points with the same x/w and y/w coordinates obscure one another.

It is usual in such systems to vary the intensity of colour within one atom by darkening it towards the circumference on the basis of the z coordinate. Some systems augment this impression of sphericity by highlighting. The simplest form of highlighting is an extension of the uniform disc treatment in which additional, brighter discs, possibly off centre, are associated with each atom. More general highlighting (Phong, 1975) is computed from four unit vectors, these being the normal to the surface, the direction to a light source, the direction to the viewer and the normalized vector sum of these last two. Intensity levels may then be set as the sum of three terms: a constant, a term proportional to the scalar product of the first two vectors (if positive) and a term proportional to a high power of the scalar product of the first and last vectors; the higher the power the glossier the surface appears to be. This final term normally adds a white term, rather than the surface colour, supposing the light source to be white.

Shadows may also be rendered to give even greater realism. In addition to the z buffer and (x, y) frame buffer a second z buffer for z' values associated with x' and y' is also required. These coordinates are then related by $x' = x + \alpha z$, $y' = y + \beta z$, $z' = z$. The second buffer is a ray buffer since x' y' are the coordinates with which an illuminating ray passing through (xyz) passes through the $z = 0$ plane, and z', stored at x', y', records the depth at which this ray encounters material. Thus any two pixels $(x_1y_1z_1)$ and $(x_2y_2z_2)$ are on the same illuminating ray if their x' and y' values are equal and the one with smaller z' shadows the other. Processing a pixel at $(x_1y_1z_1)$ therefore involves first determining its visibility on the basis of the z buffer, as before, then, whether or not it is visible, setting $z'_1 = z_1$ and considering the value of z' currently stored at $x'y'$, which we call z'_2.

If $z'_1 < z'_2$ then $x_1y_1z_1$ is in light and must be loaded accordingly. From z'_2 we find the previously processed pixel $(x_2y_2z_2)$ which is now in shade and which was in light when originally processed, so that the colour value stored at x_2y_2 needs to be altered *unless* the pixel at x_2y_2 is now $(x_2y_2z_3)$ with $z_3 < z_2$, in which case the pixel $(x_2y_2z_2)$ which has now become shadowed by $(x_1y_1z_1)$ has, in the meantime, been obscured by $(x_2y_2z_3)$ which is not shadowed by $(x_1y_1z_1)$ and no change is therefore needed. In either event z'_1 then replaces z'_2.

If $z'_1 > z'_2$ then $(x_1y_1z_1)$, if visible, is in shade and must be coloured accordingly, and in this case z'_2 is not superseded.

This shadowing scheme corresponds to illumination by a light source at infinity in picture space or, equivalently, with a z coordinate equal to that of the eye in display space. For its implementation x, y and z may be in any convenient coordinate system, *e.g.* pixel addresses, but if x and y are expressed with the range -1 to 1 and z with the range 0 to 1 corresponding to the window then they may be identified as the quantities x/w, y/w and z/w of picture space (Section 3.3.1.3.1).

If, in the notation of Section 3.3.1.3.5, the light source is placed at (P, Q, E, V) in display space and a ray leaves it in the direction (p, q, r, V) then

$$x' = \frac{p}{r} \cdot \frac{2(S-E)}{(R-L)} + \frac{2(S-E)(P-C)}{(N-E)(R-L)} + \frac{2C-R-L}{R-L},$$

which varies only with beam direction,

$$\alpha = \frac{2(S-E)(F-N)(P-C)}{(F-E)(N-E)(R-L)}$$

and similarly for y' and β.

3.3.1.5.6. Advanced hidden-line and hidden-surface algorithms

Hidden surfaces may be handled quite generally with the z-buffer technique described in the previous section but this technique becomes very inefficient with very complicated scenes. Faster techniques have been developed to handle computations in real time (*e.g.* 25 frames s^{-1}) on raster machines when both the viewpoint and parts of the environment are moving and substantial complexity is required. These techniques generally represent surfaces by a number of points in the surface, connected by lines to form panels. Many algorithms require these panels to be planar and some require them to be triangular. Of those that permit polygonal panels, most require the polygons to be convex with no re-entrant angles. Yet others are limited to cases where the objects themselves are convex. Some can handle interpenetrating surfaces, others exclude these. Some make enormous gains in efficiency if the objects in the scene are separable by the insertion of planes between them and degrade to lower efficiency if required, for example, to draw a chain. Some are especially suited to vector machines and others to raster machines, the latter capitalizing on the finite resolution of such systems. In all of these the basic entities for consideration are entire panels or edges, and in some cases vertices, point-by-point treatment of the entire surface being avoided until after all decisions are made concerning what is or is not visible.

All of these algorithms strive to derive economies from the notion of 'coherence'. The fact that, in a cine context, one frame is likely to be similar to the previous frame is referred to as 'frame coherence'. In raster scans line coherence also exists, and other kinds of coherence can also be identified. The presence of any form of coherence may enable the computation to be concerned primarily with changes in the situation, rather than with the totality of the situation so that, for example, computation is required where one edge crosses in front of another, but only trivial actions are involved so long as scan lines encounter the projections of edges in the same order.

The choice of technique from among many possibilities may even depend on the viewpoint if the scene has a statistical anisotropy. For example, the depiction of a city seen from a viewpoint near ground level involves many hidden surfaces. Distant buildings may be hidden many times over. The same scene depicted from an aerial viewpoint shows many more surfaces and fewer overlaps. This difference may swing the balance of advantage between an algorithm which sorts first on z or one which leaves that till last.

These advanced techniques have, so far, found little application in crystallography, but this may change. Ten such techniques are critically reviewed and compared by Sutherland *et al.* (1974), and three of these are described in detail by Newman & Sproull (1973).

3.3.2. Molecular modelling, problems and approaches

By R. Diamond

This section is concerned with software techniques which permit a set of atomic coordinates for a molecule to be generated *ab initio*, or to be modified, by reference to some chosen criterion, usually the electron density. Software that can change the shape of a molecule must be cognizant of the connectivity of the molecule and the bonding characteristics of atom types. It must also have means of regaining good stereochemistry if current coordinates are poor in this respect, or of performing its

manipulations in ways which conserve essential stereochemical features. Approaches to some of these problems are outlined below. Many of the issues involved, including the topics outlined in Section 3.3.1.4 above, have been excellently reviewed by Hermans (1985), though with little reference to graphical aspects, and a comprehensive treatment of modelling methods based on energies is given by Burkert & Allinger (1982).

3.3.2.1. Connectivity

It is necessary to distinguish three different kinds of connectivity, namely structural, logical and drawing connectivities. Structural connectivity consists of the specification of the chemical bonding of the molecule and, as such, is an absolute property of the molecule. Logical connectivity consists of the specification of what part or parts of a molecule are moved, and in what way, if some stereochemical feature is altered. Logical and structural connectivity are closely related and in simple cases coincide, but the distinction is apparent, for example, if the puckering of a five-membered ring is being modelled by permitting folding of the pentagon about a line connecting nonadjacent corners. This line is then a logical connection between two moving parts, but it is not a feature of the structural connectivity.

Drawing connectivity consists of a specification of the lines to be drawn to represent the molecule and often coincides with the structural connectivity. However, stylized drawings, such as those showing the α-carbon atoms of a protein, require to be drawn with lines which are features neither of the logical nor of the structural connectivity.

3.3.2.1.1. Connectivity tables

The simplest means of storing connectivity information is by means of tables in which, for each atom, a list of indices of other atoms to which it is connected is stored. This approach is quite general; it may serve any type of molecular structure and permits structures to be traversed in a variety of ways. In this form, however, it is extravagant on storage because every connection is stored twice, once at each of the nodes it connects. It may, however, provide the starting material for the algorithm of Section 3.3.1.5.2 and its generality may justify its expense.

From such a list, lists of bonds, bond angles and dihedral angles may readily be derived in which each entry points to two, three or four atoms in the atom list. Lists of these three types form the basis of procedures which adjust the shape of a molecule to reduce its estimated potential energy (Levitt & Lifson, 1969; Levitt, 1974), and of search-and-retrieval techniques (Allen *et al.,* 1979).

Katz & Levinthal (1972) discuss the explicit specification of structural connectivity in terms of a tree structure in which, for each atom, is stored a single pointer to the connected atom nearer to the root, virtual atoms being used to allow ring structures to be treated as trees. An algorithm is also presented which allows such a tree specification to be redetermined if an atom in the tree is newly chosen as the root atom or if the tree itself is modified.

Cohen *et al.* (1981) have developed methods of handling connectivity in complicated fused- and bridged-ring systems.

3.3.2.1.2. Implied connectivity

In cases where software is required to deal only with a certain class of molecule, it may be possible to exploit the characteristics of that class to define an ordering for lists of atoms such that connectivity is implied by the ordering of items in the list. Such an ordering may successfully define one of the three types of connectivity defined in Section 3.3.2.1 but it is unlikely to be able to meet the needs of all three simultaneously. It may also be at a disadvantage when required to deal with structures not part of the class for which it is designed. Within these limitations, however, it may be exceedingly efficient. Both proteins and nucleic acids are of a class which permits their logical connectivity to be specified entirely by list ordering, and the software described in Section 3.3.3.2.6 uses no connectivity tables for this purpose. The ordering rules concerned are given by Diamond (1976*b*).

Drawing connectivity needs explicit specification in such a case; this may be done using only one 16-bit integer per atom, which may be stored as part of the atom list without the need of a separate table. This integer consists of two signed bytes which act as relative pointers in the list, positive pointers implying draw-to, negative pointers implying move-to. As each atom is encountered during drawing the right byte is read and utilized, and the two bytes are swapped before proceeding. This allows up to two bonds drawn to an atom and two bonds drawn from it, four in all, with a minimum of storage (Diamond, 1984*a*).

Brandenburg *et al.* (1981) handle drawing connectivity by enlarging the molecular list with duplicate atoms such that each is connected to the next in the list, but moves and draws still need to be distinguished.

Levitt (1971) has developed a syntax for specifying structural connectivity implicitly from a list structure which is very general, though designed with biopolymers in mind, and the work of Katz & Levinthal (1972) includes something similar.

3.3.2.2. Modelling methods

Fundamental to the design of any software for molecular modelling are the choices of modelling criteria, and of parameterization. Criteria which may be adopted might include the fitting of electron density, the minimization of an energy estimate or the matching of complementary surfaces between a pair of molecules. Parameterizations which may be adopted include the use of Cartesian coordinates of atoms as independent variables, or of internal coordinates, such as dihedral angles, as independent variables with atomic positions being dependent on these. Systems designed to suit energetic criteria usually use Cartesian coordinates since all aspects of the structure, including bond lengths, must be treated as variables and be allowed to contribute to the energy estimate. Systems designed to fit a model to observed electron density, however, may adequately meet the stereochemical requirements of modelling on either parameterization, and examples of both types appear below.

Inputs to modelling systems vary widely. Systems intended for use mainly with proteins or other polymeric structures usually work with a library of monomers which the software may develop into a polymer. Systems intended for smaller molecules usually develop the molecular structure atom-by-atom rather than a residue at a time, and systems of this kind require a very general form of input. They may accept a list of atom types and coordinates if measurement and display of a known molecule is the objective, or they may accept 'sketch-pad' input in the form of a hand-drawn two-dimensional sketch of the type conventional in chemistry, if the objective is the design of a molecule. Sketch-pad input is a feature of some systems with quantum-mechanical capabilities.

3.3.2.2.1. Methods based on conformational variables

Suppose that $\mathbf{t}$ represents a vector from the current position of an atom in the model to a target position then (see Section 3.3.1.1.3), to first order, the observational equations are

$$t_{IA} = D_{IpA}\theta_p + v_{IA}$$

in which $\boldsymbol{\theta}$ represents changes to conformational variables which may include dihedral angles, bond angles, bond lengths, and parameters determining overall position and orientation of the molecule as a whole. If every such parameter is included the model acquires $3n$ degrees of freedom for n atoms, in which case

the methods of the next section are more appropriate, but if bond lengths and some or all bond angles are being treated as constants then the above equation becomes the basis of the treatment.

$$D_{IPA} = \frac{\partial r_{IA}}{\partial \theta_P} = \varepsilon_{Ijk} n_{jP} (r_{kA} - r_{kP})$$

in which $\mathbf{n}_P$ is a unit vector defining the axis of rotation for an angular variable θ_P, $\mathbf{r}_A$ and $\mathbf{r}_P$ are position vectors of the atom A and the site of the parameter P, and $\mathbf{v}_A$ represents a residual vector.

$\sigma = v_{ia} v_{ia}$ is minimized by

$$\boldsymbol{\theta} = \boldsymbol{M}^{-1}\mathbf{V}$$

in which

$$M_{PQ} = D_{iPa} D_{iQa}$$
$$V_P = D_{iPa} t_{ia}.$$

More generally, if σ represents any scalar quantity which is to be minimized, *e.g.* an energy, then

$$V_P = -\frac{1}{2}\frac{\partial \sigma}{\partial \theta_P}$$
$$M_{PQ} = \frac{1}{2}\frac{\partial^2 \sigma}{\partial \theta_P \partial \theta_Q}.$$

It is beyond the scope of this chapter to review the methods available for evaluating $\boldsymbol{\theta}$ from these equations. Difficulties may arise from two sources:

(i) Inversion of $\boldsymbol{M}$ may be difficult if $\boldsymbol{M}$ is large or ill conditioned and impossible if $\boldsymbol{M}$ is singular.

(ii) Successful evaluation of $\boldsymbol{M}^{-1}\mathbf{V}$ will not minimize σ in one step if $\mathbf{t}$ is not linearly dependent on $\boldsymbol{\theta}$ or, equivalently, $\partial^2\sigma/\partial\theta_P\partial\theta_Q$ is not constant, and substantial changes $\boldsymbol{\theta}$ are involved. Iteration is then necessary.

Difficulties of the first kind may be overcome by gradient methods, for example the conjugate gradient method without searches if $\boldsymbol{M}$ is available or with searches if it is not available, or they may be overcome by methods based on eigenvalue decompositions. If nonlinearity is serious less dependence should be placed on $\boldsymbol{M}$ and gradient methods using searches are more valuable. In this connection Diamond (1966) introduced a sliding filter technique which produced rapid convergence in extreme conditions of nonlinearity. These topics have been reviewed elsewhere (Diamond, 1981*b*, 1984*b*) and are the subject of many textbooks (Walsh, 1975; Gill *et al.*, 1981; Luenberger, 1984).

Warme *et al.* (1972) have developed a similar system using dihedral angles as variables and a variety of alternative optimization algorithms.

The modelling of flexible rings or lengths of chain with two or more fixed parts is sometimes held to be a difficulty in methods using conformational variables, although a simple two-stage solution does exist. The principle involved is the sectioning of the space of the variables into two orthogonal subspaces of which the first is used to satisfy the constraints and the second is used to perform the optimization subject to those constraints.

The algebra of the method may be outlined as follows, and is given in more detail by Diamond (1971, 1980*a*, 1981*a*). Parametric shifts $\boldsymbol{\theta}_1$ which satisfy the constraints are solutions of

$$\mathbf{V}_1 = \boldsymbol{M}_1 \boldsymbol{\theta}_1$$

in which $\mathbf{V}_1$ and $\boldsymbol{M}_1$ depend only on the target vectors, $\mathbf{t}_1$, of the atoms with constrained positions and on the corresponding derivatives. We then find a partitioned orthogonal matrix $(\boldsymbol{A}_1\boldsymbol{B}_1)$ satisfying

$$(\boldsymbol{A}_1\boldsymbol{B}_1) = (\boldsymbol{E}_\lambda \boldsymbol{E}_0)\begin{pmatrix} \boldsymbol{R}_\lambda & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_0 \end{pmatrix}$$

in which $\boldsymbol{E}_\lambda$ are the eigenvectors of $\boldsymbol{M}_1$ having positive eigenvalues, $\boldsymbol{E}_0$ are those having zero eigenvalues, and $\boldsymbol{R}_\lambda$ and $\boldsymbol{R}_0$ are arbitrary orthogonal matrices. Then

$$\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\mathbf{V}_1 = \begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\boldsymbol{M}_1(\boldsymbol{A}_1\boldsymbol{B}_1)\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\boldsymbol{\theta}_1$$
$$= \begin{pmatrix} \boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\boldsymbol{\theta}_1$$
$$\boldsymbol{A}_1^T\boldsymbol{\theta}_1 = (\boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^T\mathbf{V}_1$$

in which the matrix to be inverted is positive definite. $\boldsymbol{A}_1$, however, is rectangular so that multiplying on the left by $\boldsymbol{A}_1$ does not necessarily serve to determine $\boldsymbol{\theta}_1$, but we may write

$$\boldsymbol{\theta}_1 = (\boldsymbol{A}_1\boldsymbol{B}_1)\begin{pmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\psi}_1 \end{pmatrix}$$

giving

$$\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\mathbf{V}_1 = \begin{pmatrix} \boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\psi}_1 \end{pmatrix}$$
$$\boldsymbol{\varphi}_1 = (\boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^T\mathbf{V}_1$$

and $\boldsymbol{\psi}_1$ is indeterminate and free to adopt any value. We therefore adopt

$$\boldsymbol{\theta}_1 = \boldsymbol{A}_1\boldsymbol{\varphi}_1 = \boldsymbol{A}_1(\boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^T\mathbf{V},$$

which is the smallest vector of parametric shifts which will satisfy the constraints, and allow $\boldsymbol{\psi}_1$ to be determined by the remaining observational equations in which the target vectors, $\mathbf{t}$, are now modified to $\mathbf{t}_2$ according to

$$\mathbf{t}_2 = \mathbf{t} - \boldsymbol{D}_2\boldsymbol{\theta}_1,$$

$\boldsymbol{D}_2$ and $\mathbf{t}_2$ being the derivatives and effective target vectors for the unconstrained atoms. We then solve

$$\mathbf{V}_2 = \boldsymbol{M}_2\boldsymbol{\theta}_2$$

in which $\boldsymbol{\theta}_2$ is required to be of the form

$$\boldsymbol{\theta}_2 = \boldsymbol{B}_1\boldsymbol{\psi}_1$$

giving

$$\boldsymbol{\theta}_2 = \boldsymbol{B}_1(\boldsymbol{B}_1^T\boldsymbol{M}_2\boldsymbol{B}_1)^{-1}\boldsymbol{B}_1^T\mathbf{V}_2$$

and apply the total shifts

$$\boldsymbol{\Theta} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$$

to obtain a structure which is optimized within the restrictions imposed by the constraints.

It may happen that $\boldsymbol{B}_1^T\boldsymbol{M}_2\boldsymbol{B}_1$ is itself singular because there are insufficient data in the vector $\mathbf{t}_2$ to control the structure and the parametric shifts contained in $\boldsymbol{\theta}_2$ fully. In this event the same process may be applied again, basing the solution for $\boldsymbol{\theta}_2$ on

$$\begin{pmatrix} \boldsymbol{A}_2^T \\ \boldsymbol{B}_2^T \end{pmatrix}\boldsymbol{B}_1^T\boldsymbol{M}_2\boldsymbol{B}_1(\boldsymbol{A}_2\boldsymbol{B}_2)$$

so that the vectors in $\boldsymbol{B}_2$ represent the degrees of freedom which remain uncommitted. This method of application of constraints by subspace sectioning may be nested to any depth and is completely general.

A valid matrix $\boldsymbol{A}_1$ may be found from $\boldsymbol{M}_1$ by using the fact that the columns of $\boldsymbol{M}_1$ are all linear combinations of the columns of

$\boldsymbol{E}_\lambda$ and are void of any contribution from $\boldsymbol{E}_0$. It follows that $\boldsymbol{A}_1$ may be found by using the columns of $\boldsymbol{M}_1$ as priming vectors in the Gram–Schmidt process [Section 3.3.1.2.3 (i)] until the normalizing step involves division by zero. $\boldsymbol{A}_1$ is then complete if all the columns of $\boldsymbol{M}_1$ have been tried. $(\boldsymbol{A}_1\boldsymbol{B}_1)$ may then be completed by using arbitrary vectors as primers.

Manipulation of a ring of n atoms may be achieved by treating it as a chain of $(n+2)$ atoms [having $(n+1)$ bond lengths, n bond angles and $(n-1)$ dihedral angles] in which atom 1 is required to coincide with atom $(n+1)$ and atom 2 with $(n+2)$. $\mathbf{t}_1$ then contains two vectors, namely the lack-of-closure vectors at these points, and is typically zero. $\boldsymbol{A}_1$ is then found to have five columns corresponding to the five degrees of freedom of two points of fixed separation; $\boldsymbol{\theta}_1$ contains only zeros if the ring is initially closed, and contains ring-closure corrections if, through non-linearity or otherwise, the ring has opened. $\boldsymbol{B}_1$ contains $(p-5)$ columns if the chain of $(n+2)$ points has p variable parameters. It follows, if bond lengths and bond angles are treated as constants, that the seven-membered ring is the smallest ring which is flexible, that the six-membered ring (if it can be closed with the given bond angles) has no flexibility (though it may have discrete alternatives) and that it may be impossible to close a five-membered ring. Therefore some variation of bond angles and/or bond lengths is essential for the modelling of flexible five- and six-membered rings. Treating the ring as a chain of $(n+1)$ atoms is less satisfactory as there is then no control over the bond angle at the point of ring closure.

A useful concept for the modelling of flexible five-membered rings with near-constant bond angles is the concept of the pseudorotation angle P, and amplitude θ_m, for which the jth dihedral angle is given by

$$\theta_j = \theta_m \cos\left(P + \frac{4\pi j}{5}\right).$$

This formulation has the property $\sum_{j=0}^{4}\theta_j = 0$, which is not exactly true; nevertheless, θ_j values measured from observed conformations comply with this formulation within a degree or so (Altona & Sundaralingam, 1972).

Software specialized to the handling of condensed ring systems has been developed by van der Lieth *et al.* (1984) (Section 3.3.3.3.1) and by Cohen *et al.* (1981) (Section 3.3.3.3.2).

3.3.2.2.2. Methods based on positional coordinates

Modelling methods in which atomic coordinates are the independent variables are mathematically simpler than those using angular variables especially if the function to be minimized is a quadratic function of interatomic distances or of distances between atoms and fixed points. The method of Dodson *et al.* (1976) is representative of this class and it may be outlined as follows. If **d** is a column vector containing ideal values of the scalar distances from atoms to fixed target points or to other atoms, and if **l** is a column vector containing the prevailing values of these quantities obtained from the model, then

$$\mathbf{d} = \mathbf{l} + \boldsymbol{D}\boldsymbol{\delta}\mathbf{x} + \boldsymbol{\varepsilon}$$

in which the column matrix $\boldsymbol{\delta}\mathbf{x}$ contains alterations to the atomic coordinates, $\boldsymbol{\varepsilon}$ contains residual discrepancies and $\boldsymbol{D}$ is a large sparse rectangular matrix containing values of $\partial l/\partial x$, of which there are not more than six nonzero values on any row, consisting of direction cosines of the line of which **l** is the length. $\boldsymbol{\varepsilon}^T\boldsymbol{W}\boldsymbol{\varepsilon}$ is then minimized by setting

$$\boldsymbol{D}^T\boldsymbol{W}(\mathbf{d}-\mathbf{l}) = \boldsymbol{D}^T\boldsymbol{W}\boldsymbol{D}\boldsymbol{\delta}\mathbf{x},$$

which they solve by the method of conjugate gradients without searches. This places reliance on the linearity of the observational equations (Diamond, 1984*b*). It also works entirely with the sparse matrix $\boldsymbol{W}^{1/2}\boldsymbol{D}$, the dense matrix $\boldsymbol{D}^T\boldsymbol{W}\boldsymbol{D}$, and its inverse, being never calculated.

The method is extremely efficient in annealing a model structure for which an initial position for every atom is available, especially if the required shifts are within the quasi-linear region, but is less effective when large dihedral-angle changes are involved or when many atoms are to be placed purely by interpolation between a few others for which target positions are available. Interbond angles are controlled by assigning d values to second-nearest-neighbour distances; this is effective except for bond angles near 180° so that, in particular, planar groups require an out-of-plane dummy atom to be included which has no target position of its own but does have target values of distances between itself and atoms in the planar group. The method requires a d value to be supplied for every type of nearest- and next-nearest-neighbour distance in the structure, of which there are many, together with W values which are the inverse variances of the distances concerned as assessed by surveys of the corresponding distances in small-molecule structures, or from estimates of their accuracy, or from estimates of accuracy of the target positions.

Hermans & McQueen (1974) published a similar method which differs in that it moves only one atom at a time, in the environment of its neighbours, these being considered fixed while the central atom is under consideration. This is inefficient in the sense that in any one cycle one atom moves only a small fraction (∼3%) of the distance it will ultimately be required to move, but individual atom cycles are so cheap and simple that many cycles can be afforded. The method was selected for inclusion in *Frodo* by Jones (1978) (Section 3.3.3.2.7) and is an integral part of the *GRIP* system (Tsernoglou *et al.*, 1977; Girling *et al.*, 1980) (Section 3.3.3.2.2) for which it was designed.

3.3.2.2.3. Approaches to the problem of multiple minima

Modelling methods which operate by minimizing an objective function of the coordinates (whether conformational or positional) suffer from the fact that any realistic objective function representing the potential energy of the structure is likely to have many minima in the space of the variables for any but the simplest problems. No general system has yet been devised that can ensure that the global minimum is always found in such cases, but we indicate here two approaches to this problem.

The first approach uses dynamics to escape from potential-energy minima. Molecular-mechanics simulations allow each atom to possess momentum as well as position and integrate the equations of motion, conserving the total energy. By progressively removing energy from the simulation by scaling down the momentum vectors some potential-energy minimum may be found. Conversely, a minimization of potential energy which has led to a minimum thought not to be the global minimum may be continued by introducing atomic momenta sufficient to overcome potential-energy barriers between minima, and subsequently attenuate the momenta again. In this way a number of minima may be found (Levitt & Warshel, 1975). It is equivalent to initializing a potential-energy minimization from a number of different conformations but it has the property that the minima so found are separated by energy barriers for which an upper limit is known so that the possibility exists of exploring transition pathways.

A second approach is described by Purisima & Scheraga (1986). If the objective function to be minimized can be expressed in terms of interatomic distances, and if each atom is given coordinates in a space of $n-1$ dimensions for n atoms, then a starting structure may be postulated for which the interatomic distances *all* take their ideal values and the objective function is then necessarily at an absolute minimum. This multidimensional structure is then projected into a space of fewer dimensions, within which it is again optimized with respect to the

objective function. The dimensionality of the model is thus progressively reduced until a three-dimensional model is attained at a low energy. This means that the minimum so attained in three dimensions is approached *from beneath*, having previously possessed a lower value in a higher-dimensional space. This, in itself, does not guarantee that the three-dimensional minimum-energy structure so found is at the global minimum, but it is not affected by energy barriers between minima in the same way, and it does appear to reach very low minima, and frequently the global one. Because it is formulated entirely in terms of inter-atomic distances it offers great promise for modelling molecules on the basis of data from nuclear magnetic resonance.

3.3.3. Implementations

By R. Diamond

In this section the salient characteristics of a number of systems are described. Regrettably, it cannot claim to be a complete guide to all existing systems, but it probably describes a fairly representative sample. Some of these systems have arisen in academia and these are freely described. Some have arisen in or been adopted by companies which now market them, and these are described by reference to the original publications. Other marketed systems for which originators' published descriptions have not been found are not described. Yet other systems have been developed, for example, by companies within the chemical and pharmaceutical industries for their own use, and these have generally not been described in what follows since it is assumed that they are not generally available, even where published descriptions exist.

Software concerned especially with molecular dynamics has not been included unless it also provides static modelling capability, since this is a rapidly growing field and it has been considered to be beyond the intended scope of this chapter. Systems for which outline descriptions have already been given (Levitt & Lifson, 1969; Levitt, 1974; Diamond, 1966; Warme *et al.*, 1972; Dodson *et al.*, 1976; Hermans & McQueen, 1974) are not discussed further.

For some of the earliest work Levinthal (1966) still makes interesting reading and Feldmann (1976) is still an excellent review of the technical issues involved. The issues have not changed, the algorithms there described are still valuable, only the manner of their implementation has moved on as hardware has developed. A further review of the computer generation of illustrations has been given by Johnson (1980). Excellent bibliographies relating to these sections have been given by Morffew (1983, 1984), which together contain over 250 references including their titles.

The following material is divided into three sections. The first is concerned primarily with display rather than modelling though some of these systems can modify a model, the second is concerned with molecular modelling with reference to electron density and can develop a model *ab initio*, and the third is concerned with modelling with reference to other criteria.

Where software names are known to be acronyms constructed from initial letters, or where the original authors have used capitals, the names are capitalized here. Otherwise names are lower case with an initial capital.

While it is recognised that many of the systems here described are now of mainly historical interest, most were retained for the second edition, some were updated and some new paragraphs were added.

For the third edition a whole new section, Section 3.3.4, has been added, which provides an overview of recent developments relevant in particular to small and medium-sized structures. Recent developments in the macromolecular field of model building and computer graphics are now covered by Chapters 17.1 and 17.2 in *International Tables for Crystallography* Volume F (2001).

3.3.3.1. Systems for the display and modification of retrieved data

One of the earliest systems designed for information retrieval and display was that described by Meyer (1970, 1971) which used television raster technology and enabled the contents of the Brookhaven Protein Data Bank (Meyer, 1974; Bernstein *et al.*, 1977) to be studied visually by remote users. It also enabled a rigid two-ring molecule to be solved from packing considerations alone (Hass *et al.*, 1975; Willoughby *et al.*, 1974). Frames for display were written digitally on a disk and the display rate was synchronized to the disk rotation. With the reduction in the cost of core storage, contemporary systems use large frame buffer memories thus avoiding synchronization problems and permitting much richer detail than was possible in 1970. A majority of the systems in this section use raster techniques which preclude real-time rotation except for relatively simple drawings, though *GRAMPS* is an exception (O'Donnell & Olson, 1981; Olson, 1982) (Section 3.3.3.1.4).

3.3.3.1.1. ORTEP

This program, the Oak Ridge Thermal Ellipsoid Program, due to Johnson (1970, 1976) was developed originally for the preparation of line drawings on paper though versions have since been developed to suit raster devices with interactive capability.

The program draws molecules in correct perspective with each atom represented by an ellipsoid which is the equi-probability surface for the atomic centre, as determined by anisotropic temperature factor refinement, the principal axes of which are displayed. Bonds are represented by cylindrical rods connecting the atoms which in the drawing are tapered by the perspective.

In line-drawing versions the problem of hidden-line suppression is solved analytically, whereas the later versions for raster devices draw the furthest elements of the picture first and either overwrite these with nearer features of the scene if area painting is being done or use the nearer features as erase templates if line drawings are being made.

3.3.3.1.2. Feldmann's system

R. J. Feldmann and co-workers (Feldmann, 1983) at the National Institutes of Health, Bethesda, Maryland, USA, were among the first to develop a suite of programs to display molecular structure using colour raster-graphics techniques. Their system draws with coloured shaded spheres, usually with one sphere to represent each atom, but alternatively the spheres may represent larger moieties like amino acids or whole proteins if lower-resolution representations are required. These workers have made very effective use of colour. Conventionally, oxygens have been modelled in red, but this system allows charged oxygens to be red and uncharged ones to be pink, with a similar treatment in blue for charged and uncharged nitrogens. By such means they have been able to give immediacy to the hydrophobic and electrostatic properties of molecular surfaces, and have used these characteristics effectively in studies of the binding possibilities of benzamidine derivatives to trypsin (Feldmann *et al.*, 1978).

The algorithm developed by Porter (1978) for shading spheres to be darkened near their peripheries also computes the proper appearance of the line of intersection of two spheres wherever interpenetration occurs, in contrast to some simpler systems which draw a complete disc for whichever sphere is forward of the other. Provided that all opaque spheres are drawn first, the system is also capable of representing transparent spheres by darkening the colour of the existing background inside, and especially near the edge of, discs representing transparent foreground spheres.

Other systems that produce space-filling pictures of a similar general character have been produced by Motherwell (1978), by Sundaram & Radhakrishnan (1979) and by Lesk (next section).

3.3.3.1.3. Lesk & Hardman software

The complexity of macromolecules is a formidable obstacle to perceiving the basic features of their construction. Stylized drawings to facilitate this perception were pioneered by A. Rossmann and B. Furugren, and developed further by Richardson (1977, 1981, 1985) and by Lesk and Hardman, whose software is capable of mixing several styles of representation, among them the style of cylinders for α-helices, arrows for β-strands and ribbons for less-organized regions, or the creased-ribbon technique for the whole chain, or a ball-and-stick representation of atoms and bonds, or space-filling spheres. All these styles are available simultaneously in a single picture with depth cueing, colour and shading, and hidden-feature suppression as appropriate. It is also able to show a stylized drawing of a complete molecule together with a magnified part of it in a more detailed style. See Lesk & Hardman (1982, 1985) and Lesk (1991).

3.3.3.1.4. GRAMPS

This system, due to O'Donnell & Olson (O'Donnell & Olson, 1981; Olson, 1982) provides a high-level graphics language and its associated interpretive software. It provides a general means of defining objects, drawable by line drawings, in such a way that these may be logically connected in groups or trees using a simple command language. Such a system may, for example, define a subunit protein of an icosahedral virus and define icosahedral symmetry, in such a way that modification of one subunit is expressed simultaneously in all subunits whilst preserving the symmetry, and simultaneously allowing the entire virus particle to be rotated. Such logical and functional relationships are established by the user through the medium of the *GRAMPS* language at run time, and a great diversity of such relationships may be created. The system is thus not limited to any particular type of structure, such as linear polymers, and has proved extremely effective as a means of providing animation for the production of cine film depicting viral and other structures. *GRAMPS* runs on all Silicon Graphics workstations under IRIX 4.0 or above.

3.3.3.1.5. Takenaka & Sasada's system

Takenaka & Sasada (1980) have described a system for the manipulation and display of molecular structures, including packing environments in the crystal, using a minicomputer loosely coupled to a mainframe. Their system is also capable of model building by the addition of groups of one or more atoms with a facility for monitoring interaction distances while doing so.

3.3.3.1.6. MIDAS

This system, due to Langridge and co-workers (Gallo *et al.*, 1983; Ferrin *et al.*, 1984) is primarily concerned with the display of existing structures rather than with the establishment of new ones, but it may modify such structures by bond rotations under manual control. It is of particular value in the study of molecular interactions since two or more molecules may be manipulated simultaneously and independently. Visual docking of molecules is greatly facilitated by the display of van der Waals surfaces, which may be computed in real time so that the turning of a bond in the underlying structure does not tear the surface (Bash *et al.*, 1983).

3.3.3.1.7. Insight

This system, originally due to Dayringer *et al.* (1986), has a functionality similar to *MIDAS*. It has been replaced by *Insight* II (version 2.3.5 or later). It appears to be well suited to the study of intermolecular relationships in docking and in structural comparisons, and it is able to make modifications to structures. Objects for display may be molecular or nonmolecular, the former having an atomic substructure and the latter consisting of a vector list which may not be subdivided into referrable components. Map fitting with the current version has been reported.

3.3.3.1.8. PLUTO

PLUTO was developed by Motherwell (1978) at the Cambridge Crystallographic Data Centre (CCDC) for the display of molecular structures and crystal-packing diagrams, including an option for space-filling model style with shadowing. The emphasis was on a free format command and data structure, and the ability to produce ball-and-spoke drawings with line shadowing suitable for reproduction in journal publication. Many variant versions have been produced, with essentially the 1978 functionality, its popularity deriving from its ease of use and the provision of default options for establishing connectivity using standard bonding radii. It was distributed as part of the CCDC software associated with the Cambridge Structural Database, with an interface for reading entries from the database.

In 1993 Motherwell and others at the CCDC added an interactive menu and introduced colour and PostScript output. New features were introduced to allow interactive examination of intermolecular contacts, particularly hydrogen-bonded networks, and sections through packing diagrams (Cambridge Structural Database, 1994).

3.3.3.1.9. MDKINO

This system, due to Swanson *et al.* (1989), provides for the extraction and visualization of selected regions from molecular-dynamics simulations. It permits stereo viewing, interactive geometric interrogation and both forwards and backwards display of motion.

3.3.3.2. Molecular-modelling systems based on electron density

Systems described in this section require real-time rotation of complicated transparent scenes and all used vector-graphics technology in their original implementations for that reason, though many are now available for raster machines. In every case the graphics are the means of communication between the user and software possessing high functionality, capable of building a representation of a molecule *ab initio* and to alter it, change its shape and position it optimally in relation to an electron-density map, with due attention being paid to stereochemical considerations, by one or more of several approaches.

3.3.3.2.1. CHEMGRAF

Katz & Levinthal (1972) have developed a powerful modelling and display system for macromolecules known as *CHEMGRAF*. This system permits the definition of many atom types which includes bonding specifications, so that, for example, four types of carbon atom are included in the basic list and others may be added. A molecular fragment with an unsatisfied valency (by which it might later be attached to another such group) would have that feature represented by a 'vanishing virtual' atom which removes the need for any organizational distinction between such fragments and complete molecules. Fragments, such as amino-acid residues, may be assembled from atoms, and molecules may be assembled from atoms and/or from such fragments invoked by name, by the superposition and elimination of the relevant vanishing virtuals. The assembly process includes the development of a connectivity tree for the molecule and provision is made for the 'turning' or reconstruction of such trees if the combination of such fragments redefines the root atom of one or more of the fragments. The system also provides for ring closure. Model building initially uses fixed bond lengths and angles,

varying only the dihedral angles in single bonds, but has a library of preformed rings which could not otherwise be modelled on the simple basis. The results of such modelling may then be subjected to an energy-minimization routine using a steepest-descent method in the space of the dihedral angles and referring to the Lennard–Jones potential for nonbonded atom pairs. Atoms are first sorted into contiguous cubes so that all neighbours of any atom may be found by searching not more than 27 cubes.

The system is also capable of modelling by reference to electron density either by the translation and rotation of molecular fragments and the rotation of rotatable bonds within them or by the automatic linking of peaks in an electron-density map which are separated by less than, say, 1.8 Å, which is an important aid to interpretation when the resolution is sufficiently high.

3.3.3.2.2. GRIP

This system, developed by Professor F. P. Brooks, Dr W. V. Wright and associates at the University of North Carolina, Chapel Hill, NC, USA, was designed for biopolymers and was the first to enable a protein electron-density map to be interpreted *ab initio* without the aid of mechanical models (Tsernoglou *et al.*, 1977). Girling *et al.* (1980) give a more recent example of its use.

In its 1975 version *GRIP* is a three-machine system. Centrally there is a minicomputer which receives inputs from the user and controls a vector display with high-speed matrix-multiplication capability. The third machine is a mainframe computer with high-speed communication with the minicomputer.

The system develops a polymer chain from a library of monomers and manipulates it through bond rotations or free rotations of fragments explicitly specified by the user, with the aid of dials which may be coupled to bonds for the purpose. Bond rotations in the main chain either rotate part of the molecule relative to the remainder, which may have undesirable long-range effects, or the scope of the rotation is artificially limited with consequential discontinuities arising in the chain. Such discontinuities are removed or alleviated by the mainframe computer using the method of Hermans & McQueen (1974), which treats atomic position vectors, rather than bond rotations, as independent variables.

The system made pioneering use of a two-axis joystick to control orientation and a three-translation joystick to control position.

3.3.3.2.3. Barry, Denson & North's systems

These systems (Barry & North, 1971; North *et al.*, 1981; North, 1982) are examples of pioneering work done with minicomputers before purpose-built graphics installations became widespread; examples of their use are given by Ford *et al.* (1974), Potterton *et al.* (1983) and Dodson *et al.* (1982). They have the ability to develop a polymer chain in sections of several residues, each of which may subsequently be adjusted to remove any misfit errors where the sections overlap. Manipulations are by rotation and translation of sections and by bond rotations within sections. These movements are directly controlled by the user, who may simultaneously observe on the screen the agreement with electron density, or calculated estimates of potential or interaction energy, or a volume integral of the product of observed and model densities, or predicted shifts of proton magnetic resonance spectra. Thus models which are optimal by various criteria may be constructed, but there is no optimizer directly controlling the rotational adjustments which are determined by the user.

One of the earliest applications of them (Beddell, 1970) was in the fitting of substrate molecules to the active site of lysozyme using difference electron densities; however, the systems also permitted the enzyme–substrate interaction to be studied simultaneously and to be taken into account in adjusting the model.

3.3.3.2.4. MMS-X

The Molecular Modelling System-X (*MMS-X*) is a system of purpose-built hardware developed by Barry, Marshall and others at Washington University, St Louis. Associated with it are several sets of software. The St Louis software consists of a suite of programs rather than one large one and provides for the construction of a polymer chain in helical segments which may be adjusted bodily to fit the electron density, and internally also if the map requires this too. Nonhelical segments are built helically initially and unwound by user-controlled rotations in single bonds. The fitting is done to visual criteria. An example of the use of this system is given by Lederer *et al.* (1981).

Miller *et al.* (1981) have described an alternative software system for the same equipment. Functions invoked from a keyboard allow dials to be coupled to dihedral angles in the structure. Their system communicates with a mainframe computer which can deliver small blocks of electron density to be contoured and stored locally by the graphics system; this provides freedom of choice of contour level at run time. An example of the use of this system is given by Abad-Zapatero *et al.* (1980).

3.3.3.2.5. Texas A&M University system

This system (Morimoto & Meyer, 1976), a development of Meyer's earlier system (Section 3.3.3.1), uses vector-graphics technology and a minicomputer and is free of the timing restrictions of the earlier system. The system allows control dials to be dynamically coupled by software to rotations or translations of parts of the structure, thus permitting the reshaping or repositioning of the model to suit an electron-density map which may be contoured and managed by the minicomputer host. The system may be used to impose idealized geometry, such as planar peptides in proteins, or it may work with nonidealized coordinates.

The system was successfully applied to the structures of rubredoxin and the extracellular nuclease of *Staphylococcus aureus* (Collins *et al.*, 1975) and to binding studies of sulfonamides to carbonic anhydrase (Vedani & Meyer, 1984). In addition, two of the first proteins to be constructed without the aid of a 'Richards' Box' were modelled on this system: monoclinic lysozyme in 1976 (Hogle *et al.*, 1981) and arabinose binding protein in 1978 (Gilliland & Quiocho, 1981).

3.3.3.2.6. Bilder

This system (Diamond, 1980*b*, 1981*a*, 1982*b*) runs on a minicomputer independent of any mainframe. It builds a polymer chain from a library of residues and adapts it by internal rotations and overall positioning in much the same way as previous systems described in this section. Like them, it can provide user-controlled bond rotations, but its distinctive feature is that it has an optimizer within the minicomputer which will determine optimal combinations of bond rotations needed to meet the user's declared objectives. Such objectives are normally target positions for atoms set by the user by visual reference to the density, using the method of Section 3.3.1.3.9, but they may include target values for angles. These latter may either declare a required shape that is to take precedence over positional requirements, which are then achieved as closely as the declared shape allows, or they may be in least-squares competition with the positional requests. The optimizer also recognizes the constraints imposed by chain continuity and enables an internal section of the main chain to be modified without breaking its connection to the rest of the molecule. Similar techniques also allow ring systems to adopt various conformations, by bond rotation, without breaking the ring, simultaneously permitting the ring to have target positions. The optimizer is unperturbed by underdetermined situations, providing a minimum-disturbance result in such cases. All these properties of the optimizer are generated without recourse to any 'special cases' by a general-

ization of the subspace section technique which was used to maintain chain continuity in a 'real-space-refinement' program (Diamond, 1971). This is based entirely on the rank of the normal matrix that arises during optimization, which may serve to satisfy a constraint such as chain continuity or ring closure and simultaneously to establish what degrees of freedom remain to be controlled by other criteria. In *Bilder* this is achieved without establishing eigenvalues or eigenvectors. The method is described in outline in Section 3.3.2.2.1 and in detail by Diamond (1980*a*, 1981*a*).

The angular variables used normally comprise all single bonds but may include others, such as the peptide bond with or without a target of 180°. Thus this bond may be completely rigid, elastic, or completely free. Any interbond angles may also be parameterized but at some cost in storage. The normal mode of working is to develop a single chain for the entire length of the molecule, but if cumulative error makes fitting difficult a fresh chain may be started at any stage. *Bilder* may itself reconnect such chains at a later stage.

Construction and manipulation operates on a few residues at a time within the context of a polymer chain, but any or all of the rest of the molecule, or other molecules, may be displayed simultaneously.

Contouring is done in advance to produce a directoried file of contoured bricks of space, each brick containing up to 20 independently switchable elements which need not all be from the same map. Choice of contour level and displayed volume is thus instantaneous within the choices prepared.

The system is menu driven from a tablet, only file assignments and the like requiring the keyboard, and it offers dynamic parallax as an aid to 3D perception (Diamond *et al.*, 1982). Bloomer *et al.* (1978), Phillips (1980) and Evans *et al.* (1981) give examples of its use.

3.3.3.2.7. Frodo

This system, due to Jones (Jones, 1978, 1982, 1985; Jones & Liljas, 1984), in its original implementation was a three-machine system comprising graphics display, minicomputer and mainframe, though more recent implementations combine the last two functions in a 'midi'. Its capabilities are similar to those of *Bilder* described above, but its approach to stereochemical questions is very different. Where *Bilder* does not allow an atom to be moved out of context (unless it comprises a 'chain' of one atom) *Frodo* will permit an atom or group belonging to a chain to be moved independently of the other members of the chain and then offers regularization procedures based on the method of Hermans & McQueen (1974) to regain good stereochemistry. During this regularization, selected atoms may be fixed, remaining atoms then adjusting to these. A peptide, for example, may be inverted by moving the carbonyl oxygen across the peptide and fixing it, relying on the remaining atoms to rearrange themselves. (*Bilder* would do the equivalent operation by cutting the chain nearby, turning the peptide explicitly, reconnecting the chain and optimizing to regain chain continuity.) The *Frodo* approach is easy to use especially when large displacements of an existing structure are called for, but requires that ideal values be specified for all bond lengths, angles and fixed dihedrals since the system may need to regain such values in a distorted situation. *Bilder*, in contrast, never changes such features and so need not know their ideal values.

Frodo may work either with consecutive residues of a polymer chain, useful for initial building, or with a volume centred on a chosen position, which is ideal for adjusting interacting side chains which are close in space but remote in sequence.

In recent implementations *Frodo* can handle maps both in density grid form and in contour form and permits online contouring. It has also been developed (Jones & Liljas, 1984) to allow the automatic adjustment of the position and orientation of small rigid groupings by direct reference to electron density in the manner of Diamond (1971) but without the maintenance of chain continuity, which is subsequently reintroduced by regularization.

Horjales and Cambillau (Cambillau & Horjales, 1987; Cambillau *et al.*, 1984) have also provided a development of *Frodo* which allows the optimization of the interaction of a ligand and a substrate with both molecules being treated as flexible.

3.3.3.2.8. Guide

Brandenburg *et al.* (1981) have described a system which enables representations of macromolecules to be modified with reference to electron density. Such modifications include rotation about single bonds under manual control, or the movement with six degrees of freedom, also under manual control, of any part or parts of the molecule relative to the remainder. The latter operation may necessitate subsequent regularization of the structure if the moved and unmoved parts are chemically connected, and this is done as a separate operation on a different machine. The system also has the capability of displaying several molecules and of manually superimposing these on each other for comparison purposes.

3.3.3.2.9. HYDRA

This program, due to Hubbard (1985) (and, more recently, to Molecular Simulations) has several functional parts, referred to as 'heads', which all use the same data structure. The addition of further heads may be accomplished, knowing the data structure, without the need to know anything of the internal workings of existing heads.

The program contains extensive features for the display, analysis and modelling of molecular structure with particular emphasis on proteins. Display options include dotted surfaces, molecular skeletons, protein cartoons and a variety of van der Waals, ball-and-stick, and other raster-graphics display techniques such as ray tracing and shaded molecular surfaces. Protein analysis features include the analysis of hydrogen bonding, and of secondary and domain structure, as well as computational assessment of deviations from accepted protein structural characteristics such as abnormal main-chain or side-chain conformations and solvent exposure of hydrophobic amino acids. A full set of protein modelling facilities are provided including homology modelling and the 'docking' of substrate molecules. The program contains extensive tools for interactive modelling of structures from NMR or X-ray crystallographic data, and provides interfaces to molecular-mechanics and dynamics calculations. There are also database searching facilities to analyse and compare features of protein structure, and it is well suited to the making of cine films.

3.3.3.2.10. O

Jones *et al.* (1991) have developed a modelling system for proteins with a radically different approach to any of the foregoing, in that they begin by reducing the available electron-density map to a skeletal representation (Greer, 1974; Williams, 1982) which consists of a line running through the density close to its maximal values, this being the basis of a chain trace. Provisional α-carbon positions are also estimated at this stage. A database of known structures is then scanned for pentapeptides which may be superimposed on five successive positions in the chain trace, the best fit so found being taken to provide coordinates for the three central residues of the developing model. The process advances by three residues at each step, the first and last residues of the pentapeptide being used only to ensure that the central residues are built in a manner compatible with what precedes and follows.

The process ensures that conformations so built are free from improbable conformations, and the whole forms an adequate

starting structure for molecular-dynamics procedures, even though some imperfect geometry is to be expected where each tripeptide joins the next.

3.3.3.3. Molecular-modelling systems based on other criteria

Systems described within this section mostly have some form of energy minimization as their objective but some are purely geometrical. The optimization of molecules through empirical force fields has been reviewed by Allinger (1976), Burkert & Allinger (1982) and Boyd & Lipkowitz (1982). Some of these systems are in the academic domain, others are commercial. Most have capabilities exceeding the features referred to here and, of necessity, the list cannot be complete. No attempt at comparative evaluations is attempted or implied.

3.3.3.3.1. Molbuild, Rings, PRXBLD and MM2/MMP2

Liljefors (1983) has described a system for constructing representations of organic molecules. The system develops the molecule with plausible geometry and satisfied valencies at all stages of the development with explicit recognition of lone pairs and the various possible hybridization states. Growth is generally by substitution in which a substituent and the atom it is to replace are both nominated from the screen. The bond which is reconstructed in a substitution is generally a single bond. Double and triple bonds are introduced by the substitution of moieties containing them. Atom types may be changed after incorporation in the growing molecule, so that although the menu of substituents includes $-CH_3$ but not $-NH_2$ the latter may be obtained by incorporating $-CH_3$, then changing C to N and one of the hydrogens to a lone pair. Facilities are also provided for cyclization and acyclization.

van der Lieth *et al.* (1984) have described an extension to this that is specialized to the construction of fused-ring systems. It permits the joining of rings by fusion of a bond, in which two adjacent atoms in one ring are superposed on two in another. It also permits the construction of spiro links in which one atom is common to two rings, or the construction of bridges, or the polymerization of ring systems to form, for example, oligosaccharides. Again the satisfaction of valencies is maintained during building and the geometry of the result is governed by superposition of relevant atoms in the moieties involved.

PRXBLD is a molecular-model-building program which accepts two-dimensional molecular drawings in a manner similar to *Script* (Section 3.3.3.3.2) and constructs approximate three-dimensional coordinates from these. It is the model-building component of *SECS* (Simulation and Evaluation of Chemical Synthesis) (Wipke *et al.*, 1977; Wipke & Dyott, 1974; Wipke, 1974). See also Anderson (1984).

All three of these programs produce output which is acceptable as input to *MM*2(82)/*MMP*2 which are developments of Allinger's geometrical optimization based on molecular mechanics (Allinger, 1976).

3.3.3.3.2. Script

This system, described by Cohen *et al.* (1981), is specialized for fused-ring systems, especially steroids, but is not limited to these classes. The system allows the user to draw on the screen (with a light pen or equivalent) a two-dimensional representation of a molecule using single lines for single bonds, double lines for double bonds, and wedges to indicate out-of-plane substituents. The software can then enumerate the possible distinct conformers, each of which is expected to be near an energy minimum on the conformational potential surface. Each conformer may then be annealed to reach an energy minimum using an energy estimate based on bond lengths, bond angles, torsion angles and van der Waals, electrostatic and hydrogen-bonding terms. An example is given of the identification of an unusual conformer as the most stable one from twelve possibilities for a four-ring system.

The program is a development of similar work by Cohen (1971) in which the molecule was defined in terms of a tree structure and an optimizer based on search techniques rather than gradient vectors was used. The method included van der Waals terms and hence estimated energy differences between stereoisomers in condensed ring systems arising from steric hindrance.

3.3.3.3.3. CHARMM

This system, due to Brooks *et al.* (1983), is primarily concerned with molecular dynamics but it includes the capability of model-building proteins and nucleic acids from sequence information and values of internal coordinates (bond lengths, bond angles and dihedral angles). The resulting structure (or a given structure) may then be optimized by minimizing an empirical energy function which may include electrostatic and hydrogen-bonding terms as well as the usual van der Waals energy and a Hookean treatment of the covalent skeleton. Hydrogen atoms need not be handled explicitly, groups such as $-CH_2-$ being treated as single pseudo atoms, and this may be advisable for large structures. For small or medium proteins hydrogens may be treated explicitly and their initial positions may be determined by *CHARMM* if they are not otherwise known.

3.3.3.3.4. Commercial systems

A number of very powerful molecular-modelling systems are now available commercially and we mention a few of these here. Typically, each consists of a suite of programs sharing a common data structure so that the components of a system may be acquired selectively.

The *Chem-X* system, from Chemical Design Ltd, enables models to be developed from sketch-pad input, provides for their geometrical optimization and interfaces the result to *Gaussian*80 for quantum-mechanical calculations.

MACCS, from Molecular Design Ltd, and related software (Allinger, 1976; Wipke *et al.*, 1977; Potenzone *et al.*, 1977) has similar features and also has extensive database-maintenance facilities including data on chemical reactions.

Sybyl, from Tripos Associates (van Opdenbosch *et al.*, 1985), also builds from sketches with a standard fragment library, and provides interfaces to quantum-mechanical routines, to various databases and to *MACCS*.

Insight II (Section 3.3.3.1.7) is available from Biosym and *GRAMPS* (Section 3.3.3.1.4) is available from T. J. O'Donnell Associates.

3.3.4. Graphics software for the display of small and medium-sized molecules

By L. M. D. Cranswick

3.3.4.1. Introduction

In the age of the Internet, a wide variety of software can be easily obtained for the display of small and medium-sized molecules. An obvious question to ask is 'which software is the best?'. It can be 'best' to try all the relevant available software quickly before settling on two to three programs that are found to be most suitable. Not relying on just a single program can be important for visual and numeric cross validation of the resulting structural plot, as bugs (if present) can be quite subtle in their effect, but moderately easy to find by comparing plots made by different programs.

Most software can import crystallographic data in some format. The IUCr CIF format [see *International Tables for*

Table 3.3.4.1. *Functionality of software for crystal structure display*

Program	Ball and stick	ADPs	MSDA	Polyhedral display	Cartesian coordinates	Comparison/ overlay of multiple structures	Extended structures/ topology analysis	Magnetic structures	Incommensurate structures
ATOMS	Yes	Yes		Yes	Yes			Yes	
Balls&Sticks	Yes			Yes					
BALSAC	Yes				Yes				
Cameron	Yes	Yes							
CaRIne	Yes			Yes	Yes				
Crystallographica	Yes	Yes		Yes					
CrystalMaker	Yes			Yes	Yes			Yes	
Crystal Studio	Yes	Yes		Yes	Yes				
CrystMol	Yes	Yes			Yes	Yes			
Diamond	Yes	Yes		Yes	Yes			Yes	
DrawXTL	Yes	Yes		Yes				Yes	
FpStudio	Yes	Yes						Yes	Yes
GRETEP	Yes	Yes					Yes		
Mercury	Yes				Yes				
MolXtl	Yes	Yes			Yes				
OLEX	Yes				Yes		Yes		
ORTEP-III	Yes	Yes							
ORTEP-3 for Windows	Yes	Yes	Yes		Yes				
ORTEX/Oscail X	Yes	Yes							
PEANUT	Yes	Yes	Yes						
Platon/Pluton	Yes	Yes			Yes				
PowderCell	Yes								
PRJMS	Yes								Yes
SCHAKAL	Yes				Yes				
STRUPLO	Yes			Yes					
STRUPLO for Windows				Yes	Yes				
STRUVIR				Yes					
VENUS	Yes	Yes		Yes	Yes			Yes	
XmLmctep	Yes	Yes							
X-Seed	Yes	Yes			Yes				
Xtal-3D	Yes	Yes		Yes				Yes	
XtalDraw	Yes	Yes		Yes					

Crystallography Volume G (2005)] is slowly becoming a standard in this regard, displacing the single-crystal *SHELX* INS format, which has been a *de-facto* standard file format for much crystallographic data exchange. Entering crystallographic data by hand is slow and often introduces errors *via* typographical mistakes. Such mistakes can be minimized by importing structures using a known file type, or reformatting using a text editor or spreadsheet program into a known file type. A variety of software programs can be used for translating crystallographic structure files; however, the output (especially the handling of the symmetry operators and the space group) should be carefully checked. The CCP14 website (Cockroft & Stephenson, 2005) lists a variety of programs that can be used for this, of which a specialist program is *Cryscon* (Dowty, 2005).

3.3.4.2. Types of crystal structure display and functionality

The following information was current at the time of writing, but most software is continually changing with the insertion of new features. Thus occasional checks for updated functionality can be useful. Most software distributions include an 'updates' file containing new features and bug fixes. Detailed information on the software referred to in this section, including functionality, authorship, source and availability, is given in Tables 3.3.4.1 and 3.3.4.2.

3.3.4.2.1. Ball and stick

This is one of the most fundamental methods of displaying a crystal structure and almost all software supports this. The exceptions are *STRUVIR* and *STRUPLO for Windows* [a port of *STRUVIR* incorporating a graphical user interface (GUI)], which are both optimized for the polyhedral display of crystal structures.

3.3.4.2.2. Anisotropic displacement parameters

A subset of the programs that display ball-and-stick structures can also display surfaces related to anisotropic displacement parameters (ADPs) (also known colloquially as 'thermals', 'anisotropic thermal ellipsoids' or 'ORTEPS'). By default, most programs display the ellipsoid surfaces at a probability of 50% and normally allow this value to be changed to values between 1 and 99%. Programs that can draw ADPs include *ATOMS*, *Cameron*, *Crystallographica*, *CrystalMaker*, *Crystal Studio*, *CrystMol*, *Diamond*, *DrawXTL*, *FpStudio*, *GRETEP*, *MolXtl*, *ORTEP-III*, *ORTEP-3 for Windows*, *ORTEX*, *PEANUT*, *Platon*, *VENUS*, *XmLmctep*, *X-Seed* and *XtalDraw*.

3.3.4.2.3. Mean-square displacement amplitude

When a more thorough investigation of the ADPs would be informative (Hummel, Raselli & Bürgi, 1990), *PEANUT* can be used for plotting the mean-square displacement amplitude (MSDA), root-mean-square displacements (RMSDs) and difference surfaces. MSDA 'peanuts' can be displayed where the ADPs are non-positive-definite and the ellipsoids cannot be drawn. *ORTEP-3 for Windows* also has an option for plotting MSDAs. Care should be taken to ensure the resulting display is correct.

3.3.4.2.4. Polyhedral display

A method for understanding inorganic and intermetallic structures is the use of coordination polyhedra. The faces defined by the outer coordinated atoms generate a polyhedral object that is displayed instead of the individual atoms. This can aid in understanding the structures of polymeric inorganic materials involving both simple and complex tilt systems, and distorted

Table 3.3.4.2. *Availability of software for crystal structure display*

Program	Company or contact person	Contact and/or web address	Reference	Status
ATOMS	Shape Software	521 Hidden Valley Road, Kingsport, TN 37663, USA. http://www.shapesoftware.com/		Commercial
Balls&Sticks	Tadashi Ozawa	http://www.toycrate.org/	[1]	Freeware
BALSAC	Klaus Hermann	Fritz-Haber-Institut der MPG, Berlin. http://w3.rz-berlin.mpg.de/~hermann/hermann/balpam.html [Copyright (1991–2004) Klaus Hermann. All rights reserved.]		Commercial
Cameron	David Watkin	http://www.xtl.ox.ac.uk/crystals.html	[2]	Freeware
CaRIne	ESM Software	http://pro.wanadoo.fr/carine.crystallography/ [Copyright (1989–2005) C. Boudias and D. Monceau.]		Commercial
Crystallographica	Oxford Cryosystems Ltd	http://www.crystallographica.co.uk/	[3]	Commercial
CrystalMaker	CrystalMaker Software Ltd	5 Begbroke Science Park, Sandy Lane, Yarnton, OX5 1PF, UK. http://www.crystalmaker.co.uk/ (*CrystalMaker* 6 for Mac OS X)		Commercial
Crystal Studio	CrystalSoft & Crystal Systems Co. Ltd	PO Box 7006, Wattle Park, VIC 3128, Australia. http://www.crystalsoftcorp.com/ [Copyright (1999–2005) Crystal Systems Co., Ltd]		Commercial
CrystMol	David Duchamp	6209 Litchfield Lane, Kalamazoo, MI 49009-9159, USA. http://www.crystmol.com/		Commercial
Diamond	Crystal Impact	K. Brandenburg & H. Putz, Crystal Impact GbR, Postfach 1251, D-53002 Bonn, Germany. http://www.crystalimpact.com/		Commercial
DrawXTL	Larry Finger	http://lwfinger.net/drawxtl/	[4]	Freeware
FpStudio	Juan Rodriguez-Carvajal	http://www.ill.fr/dif/soft/fp/	[5]	Freeware
GRETEP	Jean Laugier	J. Laugier and B. Bochu, ENSP/Laboratoire des Matériaux et du Génie Physique, BP 46, 38042 Saint Martin d'Hères, France. http://www.ccp14.ac.uk/tutorial/lmgp/		Freeware
Mercury	Cambridge Crystallographic Data Centre	http://www.ccdc.cam.ac.uk/	[6]	Freeware
MolXtl	Dennis W. Bennett	http://www.uwm.edu/Dept/Chemistry/molxtl/	[7]	Freeware
OLEX	Oleg Dolomanov	http://www.ccp14.ac.uk/ccp/web-mirrors/lcells/olex_index.htm	[8]	Freeware
ORTEP-III	Carroll Johnson	http://www.ornl.gov/ortep/ortep.html	[9]	Freeware
ORTEP-3 for Windows	Louis Farrugia	http://www.chem.gla.ac.uk/~louis/software/ortep3/	[10]	Freeware
ORTEX/Oscail X	Patrick McArdle	http://www.nuigalway.ie/cryst/software.htm	[11], [12]	Freeware
PEANUT	Hans-Beat Bürgi	Laboratorium für chemische und mineralogische Kristallographie, Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland. E-mail: hans-beat.buergi@krist.unibe.ch	[13]	Freeware
Platon/Pluton	Anthony Spek	http://www.cryst.chem.uu.nl/platon/	[14], [15]	Freeware
PowderCell	Gert Nolze	http://www.ccp14.ac.uk/ccp/web-mirrors/powdcell/a_v/v_1/powder/e_cell.html	[16]	Freeware
PRJMS	Akiji Yamamoto	http://quasi.nims.go.jp/yamamoto/	[17]	Freeware
SCHAKAL	Egbert Keller	http://www.krist.uni-freiburg.de/ki/Mitarbeiter/Keller/schakal.html	[18]	Freeware
STRUPLO	Reinhadt Fischer	http://www.brass.uni-bremen.de/	[19]	Freeware
STRUPLO for Windows	Louis Farrugia	http://www.chem.gla.ac.uk/~louis/software/struplo/	[19], [20]	Freeware
STRUVIR	Armel Le Bail	http://www.cristal.org/vrml/struvir.html	[19], [20]	Freeware
VENUS	Fujio Izumi	http://homepage.mac.com/fujioizumi/visualization/VENUS.html	[21], [22]	Freeware
XmLmctep	Alain Soyer	http://www.lmcp.jussieu.fr/~soyer/Lmctep_en.html	[23]	Freeware
X-Seed	Len Barbour	http://x-seed.net/	[24], [25]	Commercial
Xtal-3D	Alan Hewat	http://barns.ill.fr/dif/xtal-3d.html	[26]	Freeware
XtalDraw	Bob Downs	http://www.geo.arizona.edu/xtal/xtaldraw/xtaldraw.html	[27]	Freeware

References: [1] Ozawa & Kang (2004); [2] Watkin *et al.* (1996); [3] Siegrist (1997); [4] Finger *et al.* (2007); [5] Chapon & Rodriguez-Carvajal (2005); [6] Bruno *et al.* (2002); [7] Bennett (2004); [8] Dolomanov *et al.* (2003); [9] Burnett & Johnson (1996); [10] Farrugia (1997); [11] McArdle (1994); [12] McArdle *et al.* (2004); [13] Hummel, Hauser & Bürgi (1990); [14] Spek (1998); [15] Spek (2003); [16] Kraus & Nolze (1996); [17] Yamamoto (1982); [18] Keller (1999); [19] Fischer (1985); [20] Le Bail (1996); [21] Izumi & Dilanian (2002); [22] Izumi (2004); [23] Soyer (1993); [24] Barbour (2001); [25] Atwood & Barbour (2003); [26] Hewat (2002); [27] Downs & Hall-Wallace (2003).

coordination structures. Polyhedral display is provided by the programs *ATOMS*, *Balls&Sticks*, *CaRIne*, *Crystallographica*, *CrystalMaker*, *Crystal Studio*, *Diamond*, *DrawXTL*, *STRUPLO*, *STRUPLO for Windows* (a port of *STRUVIR* incorporating a GUI), *STRUVIR*, *VENUS*, *Xtal-3D* and *XtalDraw*. Nearly all the programs display polyhedra automatically after the user has defined (i) a central atom, (ii) the coordinated atoms and (iii) minimum (often a default of near zero) and maximum bond distances. One exception is *Balls&Sticks*, where graphical point and click of the mouse is used to define the polyhedra.

3.3.4.2.5. Cartesian coordinates

Importing Cartesian coordinates can allow the display of incommensurate and quasicrystal structures if the refinement software has this as an output option. Using Cartesian coordinates can sometimes be more convenient for the slight modification of structures for the display of distortions or individual molecules. A structure defined as triclinic with space group *P*1 and a cubic cell with edges of unit length would also work for importing a structure or molecule originally defined in a Cartesian frame of reference.

3.3.4.2.6. Comparing or overlaying crystal structures

The graphical comparison of crystallographic structures can be useful and time-saving for comparison of polymorphs or a chemically similar series of small-molecule structures. One program that can perform this function is *CrystMol*, where multiple molecular structures can be compared using a point and click menu or *via* the *CrystMol* scripting system. RMS differences are also listed. Superposition of structures is discussed in Section 3.3.1.2.2.

3.3.4.2.7. Extended structures and topology analysis

Currently, the only available program that rigorously analyses extended structures (involving overlapping or interpenetrating molecules) is *OLEX*. For graphical viewing of extended structures *OLEX* displays particular fragments in a single colour. *GRETEP* also has this display functionality, making it useful for viewing extended structures.

3.3.4.2.8. Magnetic crystal structure display

The software listed in Tables 3.3.4.1 and 3.3.4.2 includes programs that can display graphics representing magnetic vectors without necessarily having the ability to understand magnetic symmetry. Programs that can display magnetic structures include *ATOMS*, *CrystalMaker*, *Diamond*, *DrawXTL*, *FpStudio*, *VENUS* and *Xtal-3D*.

3.3.4.2.9. Incommensurate crystal structures

PRJMS and *FpStudio* are currently the only programs that can plot modulated structures in three-dimensional space; *FpStudio* is currently restricted to incommensurate magnetic structures. However, importing Cartesian coordinates can be used to display incommensurate structures when incommensurate refinement software can output coordinates in this format. *BALSAC* provides a good example of software which by default uses Cartesian coordinates, from which plots of incommensurate and quasicrystal structures can be generated.

References

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Structure of southern bean mosaic virus at 2.8 Å resolution. Nature (London)*, **286**, 33–39.

Abi-Ezzi, S. S. & Bunshaft, A. J. (1986). *An implementer's view of PHIGS. IEEE Comput. Graphics Appl.* Vol. 6, Part 2.

Aharonov, Y., Farach, H. A. & Poole, C. P. (1977). *Non-linear vector product to describe rotations. Am. J. Phys.* **45**, 451–454.

Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. Acta Cryst.* B**35**, 2331–2339.

Allinger, N. L. (1976). *Calculation of molecular structure and energy by force field methods. Adv. Phys. Org. Chem.* **13**, 1–82.

Altona, C. & Sundaralingam, M. (1972). *Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. J. Am. Chem. Soc.* **94**, 8205–8212.

American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Graphical Kernel System (GKS) Functional Description (1985). ISO 7942, ISO Central Secretariat, Geneva, Switzerland.

American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Programmer's Hierarchical Graphics System (PHIGS) Functional Description, Archive File Format, Clear-Text Encoding of Archive File (1988). ANSI X3.144–1988. ANSI, New York, USA.

Anderson, S. (1984). *Graphical representation of molecules and substructure-search queries in MACCS. J. Mol. Graphics*, **2**, 83–90.

Arnold, D. B. & Bono, P. R. (1988). *CGM and CGI: Metafile and Interface Standards for Computer Graphics.* Berlin: Springer-Verlag.

Atwood, J. L. & Barbour, L. J. (2003). *Molecular graphics: from science to art. Cryst. Growth Des.* **3**, 3.

Barbour, L. J. (2001). *X-Seed – A software tool for supramolecular crystallography. J. Supramol. Chem.* **1**, 189.

Barry, C. D. & North, A. C. T. (1971). *The use of a computer-controlled display system in the study of molecular conformations. Cold Spring Harbour Symp. Quant. Biol.* **36**, 577–584.

Bash, P. A., Pattabiraman, N., Huang, C., Ferrin, T. E. & Langridge, R. (1983). *Van der Waals surfaces in molecular modelling: implementation with real-time computer graphics. Science*, **222**, 1325–1327.

Beddell, C. J. (1970). *An X-ray Crystallographic Study of the Activity of Lysozyme.* DPhil thesis, University of Oxford, England.

Bennett, D. W. (2004). *MolXtl: molecular graphics for small-molecule crystallography. J. Appl. Cryst.* **37**, 1038.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 535–542.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Boyd, D. B. & Lipkowitz, K. B. (1982). *Molecular mechanics, the method and its underlying philosophy. J. Chem. Educ.* **59**, 269–274.

Brandenburg, N. P., Dempsey, S., Dijkstra, B. W., Lijk, L. J. & Hol, W. G. J. (1981). *An interactive graphics system for comparing and model building of macromolecules. J. Appl. Cryst.* **14**, 274–279.

Brooks, B. R., Brucolleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem.* **4**, 187–217.

Brown, M. D. (1985). *Understanding PHIGS.* Template, Megatek Corp., San Diego, California, USA.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *New software for searching the Cambridge Structural Database and visualizing crystal structures. Acta Cryst.* B**58**, 389–397.

Burkert, U. & Allinger, N. L. (1982). *Molecular Mechanics. ACS Monogr.* No. 177.

Burnett, M. N. & Johnson, C. K. (1996). *ORTEP-III: Oak Ridge thermal ellipsoid plot program for crystal structure illustrations*. Report ORNL-6895. Oak Ridge National Laboratory, Tennessee, USA.

Cambillau, C. & Horjales, E. (1987). *TOM: a FRODO subpackage for protein-ligand fitting with interactive energy minimization. J. Mol. Graphics*, **5**, 174–177.

Cambillau, C., Horjales, E. & Jones, T. A. (1984). *TOM, a display program for fitting ligands into protein receptors and performing interactive energy minimization. J. Mol. Graphics*, **2**, 53–54.

Cambridge Structural Database (1994). Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

Chapon, L. C. & Rodriguez-Carvajal, J. (2005). *FpStudio*. Rutherford Appleton Laboratory, UK, and Laboratoire Léon Brillouin, Saclay, France.

Cockcroft, J. K. & Stephenson, R. (2005). *EPSRC-funded Collaborative Computational Project Number 14 for Single Crystal and Powder Diffraction (CCP14)*, http://www.ccp14.ac.uk/.

Cockrell, P. R. (1983). *A new general purpose method for large volume production of contour charts. Comput. Graphics Forum*, **2**, 35–47.

Cohen, N. C. (1971). *GEMO: a computer program for the calculation of the preferred conformations of organic molecules. Tetrahedron*, **27**, 789–797.

Cohen, N. C., Colin, P. & Lemoine, G. (1981). *Script: interactive molecular geometrical treatments on the basis of computer-drawn chemical formula. Tetrahedron*, **37**, 1711–1721.

Collins, D. M., Cotton, F. A., Hazen, E. E., Meyer, E. F. & Morimoto, C. N. (1975). *Protein crystal structures: quicker, cheaper approaches. Science*, **190**, 1047–1053.

Connolly, M. L. (1983*a*). *Solvent-accessible surfaces of proteins and nucleic acids. Science*, **221**, 709–713.

Connolly, M. L. (1983*b*). *Analytical molecular surface calculation. J. Appl. Cryst.* **16**, 548–558.

Dam, A. van (1988). *PHIGS+ functional description, revision 3.0. Comput. Graphics*, **22**, 125–218.

Dayringer, H. E., Tramontano, A., Sprang, S. R. & Fletterick, R. J. (1986). *Interactive program for visualization and modelling of proteins, nucleic acids and small molecules. J. Mol. Graphics*, **4**, 82–87.

Diamond, R. (1966). *A mathematical model-building procedure for proteins. Acta Cryst.* **21**, 253–266.

Diamond, R. (1971). *A real-space refinement procedure for proteins. Acta Cryst.* A**27**, 436–452.

Diamond, R. (1976*a*). *On the comparison of conformations using linear and quadratic transformations. Acta Cryst.* A**32**, 1–10.

Diamond, R. (1976*b*). *Model building techniques for macromolecules.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 336–343. Copenhagen: Munksgaard.

Diamond, R. (1980*a*). *Some problems in macromolecular map interpretation.* In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 21.01–21.19. Bangalore: Indian Academy of Sciences for the International Union of Crystallography.

Diamond, R. (1980*b*). *Inter-active graphics.* In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 27.01–27.16. Bangalore: Indian Academy of Sciences for the International Union of Crystallography.

Diamond, R. (1981*a*). *BILDER: a computer graphics program for biopolymers and its application to the interpretation of the structure of tobacco mosaic virus protein discs at 2.8 Å resolution.* In *Biomolecular Structure, Conformation, Function and Evolution*, Vol. 1, edited by R. Srinivasan, pp. 567–588. Oxford: Pergamon Press.

Diamond, R. (1981*b*). *A review of the principles and properties of the method of least squares.* In *Structural Aspects of Biomolecules*, edited by R. Srinivasan & V. Pattabhi, pp. 81–122. Delhi: Macmillan India Ltd.

Diamond, R. (1982*a*). *Two contouring algorithms.* In *Computational Crystallography*, edited by D. Sayre, pp. 266–272. Oxford University Press.

Diamond, R. (1982*b*). *BILDER: an interactive graphics program for biopolymers.* In *Computational Crystallography*, edited by D. Sayre, pp. 318–325. Oxford University Press.

Diamond, R. (1984*a*). *Applications of computer graphics in molecular biology. Comput. Graphics Forum*, **3**, 3–11.

Diamond, R. (1984*b*). *Least squares and related optimisation techniques.* In *Methods and Applications in Crystallographic Computing*, edited by S. R. Hall & T. Ashida, pp. 174–192. Oxford University Press.

Diamond, R. (1988). *A note on the rotational superposition problem. Acta Cryst.* A**44**, 211–216.

Diamond, R. (1989). *A comparison of three recently published methods for superimposing vector sets by pure rotation. Acta Cryst.* A**45**, 657.

Diamond, R. (1990*a*). *On the factorisation of rotations with special reference to diffractometry. Proc. R. Soc. London Ser. A*, **428**, 451–472.

Diamond, R. (1990*b*). *Chirality in rotational superposition. Acta Cryst.* A**46**, 423.

Diamond, R. (1992). *On the multiple simultaneous superposition of molecular structures by rigid body transformations. Protein Sci.* **1**, 1279–1287.

Diamond, R. (1995). *Coordinate based cluster analysis. Acta Cryst.* D**51**, 127–135.

Diamond, R., Wynn, A., Thomsen, K. & Turner, J. (1982). *Three-dimensional perception for one-eyed guys, or, the use of dynamic parallax.* In *Computational Crystallography*, edited by D. Sayre, pp. 286–293. Oxford University Press.

Dodson, E. J., Isaacs, N. W. & Rollett, J. S. (1976). *A method for fitting satisfactory models to sets of atomic positions in protein structure refinements. Acta Cryst.* A**32**, 311–315.

Dodson, G. G., Eliopoulos, E. E., Isaacs, N. W., McCall, M. J., Niall, H. D. & North, A. C. T. (1982). *Rat relaxin: insulin-like fold predicts a likely receptor binding region. Int. J. Biol. Macromol.* **4**, 399–405.

Dolomanov, O. V., Blake, A. J., Champness, N. R. & Schröder, M. (2003). *OLEX: new software for visualization and analysis of extended crystal structures. J. Appl. Cryst.* **36**, 1283–1284.

Downs, R. T. & Hall-Wallace, M. (2003). *The American Mineralogist Crystal Structure Database. Am. Mineral.* **88**, 247–250.

Dowty, E. (2005). *Cryscon*. Shape Software, 521 Hidden Valley Road, Kingsport, TN 37663, USA. http://www.shapesoftware.com/.

Enderle, G., Kansy, K. & Pfaff, G. (1984). *Computer Graphics Programming, GKS – the Graphics Standard.* Berlin: Springer-Verlag.

Evans, P. R., Farrants, G. W. & Hudson, P. J. (1981). *Phosphofructokinase: structure and control. Philos. Trans. R. Soc. London Ser. B*, **293**, 53–62.

Farrugia, L. J. (1997). *ORTEP-3 for Windows – a version of ORTEP-III with a Graphical User Interface (GUI). J. Appl. Cryst.* **30**, 565.

Feldmann, R. J. (1976). *The design of computing systems for molecular modeling. Annu. Rev. Biophys. Bioeng.* **5**, 477–510.

Feldmann, R. J. (1983). *Directions in macromolecular structure representation and display.* In *Computer Applications in Chemistry*, edited by S. R. Heller & R. Potenzone Jr, pp. 9–18. Amsterdam: Elsevier.

Feldmann, R. J., Bing, D. H., Furie, B. C. & Furie, B. (1978). *Interactive computer surface graphics approach to the study of the active site of bovine trypsin. Proc. Natl Acad. Sci. Biochemistry*, **75**, 5409–5412.

Ferrin, T. E., Huang, C., Jarvis, L. & Langridge, R. (1984). *Molecular inter-active display and simulation: MIDAS. J. Mol. Graphics*, **2**, 55.

Finger, L. W., Kroeker, M. & Toby, B. H. (2007). *DRAWxtl, an open-source computer program to produce crystal structure drawings. J. Appl. Cryst.* **40**, 188–192.

Fischer, R. X. (1985). *STRUPLO84, a Fortran plot program for crystal structure illustrations in polyhedral representation. J. Appl. Cryst.* **18**, 258–262.

Foley, J. D., van Dam, A., Feiner, S. K. & Hughes, J. F. (1990). *Computer Graphics Principles and Practice*, 2nd ed. New York: Addison Wesley.

Ford, L. O., Johnson, L. N., Machin, P. A., Phillips, D. C. & Tjian, R. (1974). *Crystal structure of a lysozyme-tetrasaccharide lactone complex. J. Mol. Biol.* **88**, 349–371.

Gallo, L., Huang, C. & Ferrin, T. (1983). *UCSF MIDAS, molecular interactive display and simulation, users' guide.* Computer Graphics Laboratory, School of Pharmacy, University of California, San Francisco, USA.

Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimization.* Orlando: Academic Press.

Gilliland, G. L. & Quiocho, F. A. (1981). *Structure of the* L*-arabinose-binding protein from Escherichia coli at 2.4 Å resolution. J. Mol. Biol.* **146**, 341–362.

Girling, R. L., Houston, T. E., Schmidt, W. C. Jr & Amma, E. L. (1980). *Macromolecular structure refinement by restrained least-squares and interactive graphics as applied to sickling deer type III hemoglobin. Acta Cryst.* A**36**, 43–50.

Gossling, T. H. (1967). *Two methods of presentation of electron-density maps using a small-store computer. Acta Cryst.* **22**, 465–468.

Greer, J. (1974). *Three-dimensional pattern recognition: an approach to automated interpretation of electron density maps of proteins. J. Mol. Biol.* **82**, 279–302.

Harris, M. R., Geddes, A. J. & North, A. C. T. (1985). *A liquid crystal stereo-viewer for molecular graphics. J. Mol. Graphics*, **3**, 121–122.

Hass, B. S., Willoughby, T. V., Morimoto, C. N., Cullen, D. L. & Meyer, E. F. (1975). *The solution of the structure of spirodienone by visual packing analysis. Acta Cryst.* B**31**, 1225–1229.

Heap, B. R. & Pink, M. G. (1969). *Three contouring algorithms*, DNAM Rep. 81. National Physical Laboratory, Teddington, England.

Hermans, J. (1985). *Rationalization of molecular models.* In *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Molecules*, Part B, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 171–189. Orlando: Academic Press.

Hermans, J. & McQueen, J. E. (1974). *Computer manipulation of (macro) molecules with the method of local change. Acta Cryst.* A**30**, 730–739.

Hewat, A. (2002). *Databases linked to electronic publications. Acta Cryst.* A**58** (Suppl.), C216.

Hogle, J., Rao, S. T., Mallikarjunan, M., Beddell, C., McMullan, R. K. & Sundaralingam, M. (1981). *Studies of monoclinic hen egg white lysozyme. Structure solution at 4 Å resolution and molecular-packing*

comparisons with tetragonal and triclinic lysozymes. Acta Cryst. B**37**, 591–597.

Hopgood, F. R. A., Duce, D. A., Gallop, J. R. & Sutcliffe, D. C. (1986). *Introduction to the Graphical Kernel System*, 2nd ed. London: Academic Press.

Hubbard, R. E. (1983). *Colour molecular graphics on a microcomputer. J. Mol. Graphics*, **1**, 13–16, C3–C4.

Hubbard, R. E. (1985). *The representation of protein structure.* In *Computer Aided Molecular Design*, pp. 99–106. Proceedings of a two-day conference, London, October 1984. London: Oyez Scientific.

Hummel, W., Hauser, J. & Bürgi, H.-B. (1990). *PEANUT: computer graphics program to represent atomic displacement parameters. J. Mol. Graphics*, **8**, 214–220.

Hummel, W., Raselli, A. & Bürgi, H.-B. (1990). *Analysis of atomic displacement parameters and molecular motion in crystals. Acta Cryst.* B**46**, 683–692.

International Standards Organisation, International Standard Information Processing Systems – Computer Graphics – Graphical Kernel System for Three Dimensions (GKS-3D), Functional Description (1988). ISO Document No. 8805:1988(E). American National Standards Institute, New York, USA.

International Tables for Crystallography (2001). Vol. F. *Crystallography of Biological Macromolecules*, edited by M. G. Rossmann & E. Arnold. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (2005). Vol. A. *Space-Group Symmetry*, edited by Th. Hahn. Heidelberg: Springer.

International Tables for Crystallography (2005). Vol. G. *Definition and Exchange of Crystallographic Data*, edited by S. R. Hall & B. McMahon. Heidelberg: Springer.

IUPAC–IUB Commission on Biochemical Nomenclature (1970). *Abbreviations and symbols for the description of the conformation of polypeptide chains. J. Biol. Chem.* **245**, 6489–6497.

Izumi, F. (2004). *Beyond the ability of Rietveld analysis: MEM-based pattern fitting. Solid State Ionics*, **172**, 1–6.

Izumi, F. & Dilanian, R. A. (2002). *Structure refinement based on the maximum-entropy method from powder diffraction data.* In *Recent Research Developments in Physics*, Vol. 3, Part II, pp. 699–726. Trivandrum: Transworld Research Network.

Johnson, C. K. (1970). *Drawing crystal structures by computer.* In *Crystallographic Computing*, edited by F. R. Ahmed, pp. 227–230. Copenhagen: Munksgaard.

Johnson, C. K. (1976). *ORTEP-II. A Fortran thermal-ellipsoid plot program for crystal structure illustrations.* Report ORNL-5138. Oak Ridge National Laboratory, Tennessee, USA.

Johnson, C. K. (1980). *Computer-generated illustrations.* In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 26.01–26.10. Bangalore: Indian Academy of Sciences for the International Union of Crystallography.

Jones, T. A. (1978). *A graphics model building and refinement system for macromolecules. J. Appl. Cryst.* **11**, 268–272.

Jones, T. A. (1982). *FRODO: a graphics fitting program for macromolecules.* In *Computational Crystallography*, edited by D. Sayre, pp. 303–317. Oxford University Press.

Jones, T. A. (1985). *Interactive computer graphics: FRODO.* In *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Molecules*, Part B, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 157–171. Orlando: Academic Press.

Jones, T. A. & Liljas, L. (1984). *Crystallographic refinement of macromolecules having non-crystallographic symmetry. Acta Cryst.* A**40**, 50–57.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst.* A**47**, 110–119.

Kabsch, W. (1976). *A solution for the best rotation to relate two sets of vectors. Acta Cryst.* A**32**, 922–923.

Kabsch, W. (1978). *A discussion of the solution for the best rotation to relate two sets of vectors. Acta Cryst.* A**34**, 827–828.

Katz, L. & Levinthal, C. (1972). *Interactive computer graphics and representation of complex biological structures. Annu. Rev. Biophys. Bioeng.* **1**, 465–504.

Kearsley, S. K. (1989). *On the orthogonal transformation used for structural comparisons. Acta Cryst.* A**45**, 208–210.

Kearsley, S. K. (1990). *An algorithm for the simultaneous superposition of a structural series. J. Comput. Chem.* **11**, 1187–1192.

Keller, E. (1999). *SCHAKAL 99, a computer program for the graphic representation of molecular and solid-state structure models*. University of Freiburg, Germany.

Kraus, W. & Nolze, G. (1996). *POWDER CELL – a program for the representation and manipulation of crystal structures and calculation of the resulting X-ray powder patterns. J. Appl. Cryst.* **29**, 301–303.

Langridge, R., Ferrin, T. E., Kuntz, I. D. & Connolly, M. L. (1981). *Real-time color graphics in studies of molecular interactions. Science*, **211**, 661–666.

Le Bail, A. (1996). *VRML as a tool for exploring complex structures. Acta Cryst.* A**52** (Suppl.), C78.

Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D. & Mathews, F. S. (1981). *Improvement of the 2.5 Å resolution model of cytochrome b_{562} by re-determining the primary structure and using molecular graphics. J. Mol. Biol.* **148**, 427–448.

Lesk, A. M. (1991). *Protein Architecture: A Practical Approach.* Oxford: IRL Press.

Lesk, A. M. & Hardman, K. D. (1982). *Computer-generated schematic diagrams of protein structures. Science*, **216**, 539–540.

Lesk, A. M. & Hardman, K. D. (1985). *Computer-generated pictures of proteins.* In *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Molecules*, Part B, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 381–390. Orlando: Academic Press.

Levinthal, C. (1966). *Molecular model-building by computer. Sci. Am.* **214**, 42–52.

Levitt, M. (1971). PhD Dissertation, ch. 2. University of Cambridge, England.

Levitt, M. (1974). *Energy refinement of hen egg-white lysozyme. J. Mol. Biol.* **82**, 393–420.

Levitt, M. & Lifson, S. (1969). *Refinement of protein conformations using a macromolecular energy minimization procedure. J. Mol. Biol.* **46**, 269–279.

Levitt, M. & Warshel, A. (1975). *Computer simulation of protein folding. Nature (London)*, **253**, 694–698.

Lieth, C. W. van der, Carter, R. E., Dolata, D. P. & Liljefors, T. (1984). *RINGS – a general program to build ring systems. J. Mol. Graphics*, **2**, 117–123.

Liljefors, T. (1983). *MOLBUILD – an interactive computer graphics interface to molecular mechanics. J. Mol. Graphics*, **1**, 111–117.

Luenberger, D. G. (1984). *Linear and Nonlinear Programming.* Reading: Addison Wesley.

McArdle, P. (1994). *ORTEX2.1 – a 1677-atom version of ORTEP with automatic cell outline and cell packing for use on a PC. J. Appl. Cryst.* **27**, 438–439.

McArdle, P., Gilligan, K., Cunningham, D., Dark, R. & Mahon, M. (2004). *A method for the prediction of the crystal structure of ionic organic compounds? The crystal structures of o-toluidinium chloride and bromide and polymorphism of bicifadine hydrochloride. CrystEngComm*, **6**, 303.

Mackay, A. L. (1984). *Quaternion transformation of molecular orientation. Acta Cryst.* A**40**, 165–166.

McLachlan, A. D. (1972). *A mathematical procedure for superimposing atomic coordinates of proteins. Acta Cryst.* A**28**, 656–657.

McLachlan, A. D. (1979). *Gene duplications in the structural evolution of chymotrypsin. Appendix: Least squares fitting of two structures. J. Mol. Biol.* **128**, 49–79.

McLachlan, A. D. (1982). *Rapid comparison of protein structures. Acta Cryst.* A**38**, 871–873.

Max, N. L. (1984). *Computer representation of molecular surfaces. J. Mol. Graphics*, **2**, 8–13, C2–C4.

Meyer, E. F. (1970). *Three-dimensional graphical models of molecules and a time-slicing computer. J. Appl. Cryst.* **3**, 392–395.

Meyer, E. F. (1971). *Interactive computer display for the three dimensional study of macromolecular structures. Nature (London)*, **232**, 255–257.

Meyer, E. F. (1974). *Storage and retrieval of macromolecular structural data. Biopolymers*, **13**, 419–422.

Miller, J. R., Abdel-Meguid, S. S., Rossmann, M. G. & Anderson, D. C. (1981). *A computer graphics system for the building of macromolecular models into electron density maps. J. Appl. Cryst.* **14**, 94–100.

Morffew, A. J. (1983). *Bibliography for molecular graphics. J. Mol. Graphics*, **1**, 17–23.

Morffew, A. J. (1984). *Bibliography for molecular graphics, 1983/84. J. Mol. Graphics*, **2**, 124–128.

Morimoto, C. N. & Meyer, E. F. (1976). *Information retrieval, computer graphics, and remote computing.* In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 488–496. Copenhagen: Munksgaard.

Motherwell, W. D. S. (1978). *Pluto – a program for displaying molecular and crystal structures.* Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

Newman, W. M. & Sproull, R. F. (1973). *Principles of Inter-active Computer Graphics.* New York: McGraw-Hill.

North, A. C. T. (1982). *Use of interactive computer graphics in studying molecular structures and interactions. Chem. Ind.* pp. 221–225.

North, A. C. T., Denson, A. K., Evans, A. C., Ford, L. O. & Willoughby, T. V. (1981). *The use of an interactive computer graphics system in the study of protein conformations.* In *Biomolecular Structure, Conformation, Function and Evolution*, Vol. 1, edited by R. Srinivasan, pp. 59–72. Oxford: Pergamon Press.

O'Donnell, T. J. & Olson, A. J. (1981). *GRAMPS – a graphics language interpreter for real-time, interactive, three-dimensional picture editing and animation. Comput. Graphics*, **15**, 133–142.

Olson, A. J. (1982). *GRAMPS: a high level graphics interpreter for expanding graphics utilization.* In *Computational Crystallography*, edited by D. Sayre, pp. 326–336. Oxford University Press.

Opdenbosch, N. van, Cramer, R. III & Giarrusso, F. F. (1985). *Sybyl, the integrated molecular modelling system. J. Mol. Graphics*, **3**, 110–111.

Ozawa, T. C. & Kang, S. J. (2004). *Balls&Sticks: easy-to-use structure visualization and animation program. J. Appl. Cryst.* **37**, 679.

Pearl, L. H. & Honegger, A. (1983). *Generation of molecular surfaces for graphic display. J. Mol. Graphics*, **1**, 9–12, C2.

Phillips, S. E. V. (1980). *Structure and refinement of oxymyoglobin at 1.6 Å resolution. J. Mol. Biol.* **142**, 531–554.

Phong, B. T. (1975). *Illumination for computer generated images. Commun. ACM*, **18**, 311–317.

Porter, T. K. (1978). *Spherical shading. Comput. Graphics*, **12**, 282–285.

Potenzone, R., Cavicchi, E., Weintraub, H. J. R. & Hopfinger, A. J. (1977). *Molecular mechanics and the CAMSEQ processor. Comput. Chem.* **1**, 187–194.

Potterton, E. A., Geddes, A. J. & North, A. C. T. (1983). *Attempts to design inhibitors of dihydrofolate reductase using interactive computer graphics with real time energy calculations.* In *Chemistry and Biology of Pteridines*, edited by J. A. Blair, pp. 299–303. Berlin, New York: Walter de Gruyter.

Purisima, E. O. & Scheraga, H. A. (1986). *An approach to the multiple-minima problem by relaxing dimensionality. Proc. Natl Acad. Sci. USA*, **83**, 2782–2786.

Richardson, J. S. (1977). *β-Sheet topology and the relatedness of proteins. Nature (London)*, **268**, 495–500.

Richardson, J. S. (1981). *The anatomy and taxonomy of protein structure. Adv. Protein Chem.* **34**, 167–339.

Richardson, J. S. (1985). *Schematic drawings of protein structures.* In *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Molecules*, Part B, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 359–380. Orlando: Academic Press.

Shapiro, A., Botha, J. D., Pastore, A. & Lesk, A. M. (1992). *A method for multiple superposition of structures. Acta Cryst.* A**48**, 11–14.

Siegrist, T. (1997). *Crystallographica – a software toolkit for crystallography. J. Appl. Cryst.* **30**, 418–419.

Soyer, A. (1993). *LMCTEP: software for crystal structure representation. J. Appl. Cryst.* **26**, 495.

Spek, A. L. (1998). *PLATON. A multipurpose crystallographic tool*. University of Utrecht, Utrecht, The Netherlands.

Spek, A. L. (2003). *Single-crystal structure validation with the program PLATON. J. Appl. Cryst.* **36**, 7–13.

Sundaram, K. & Radhakrishnan, R. (1979). *A computer program for topographic analysis of biomolecular systems. Comput. Programs Biomed.* **10**, 34–42.

Sutcliffe, D. C. (1980). *Contouring over rectangular and skewed rectangular grids – an introduction.* In *Mathematical Methods in Computer Graphics and Design*, edited by K. W. Brodie, pp. 39–62. London: Academic Press.

Sutherland, I. E., Sproull, R. F. & Schumacker, R. A. (1974). *A characterization of ten hidden surface algorithms. Comput. Surv.* **6**, 1–55.

Swanson, S. M., Wesolowski, T., Geller, M. & Meyer, E. F. (1989). *Animation: a useful tool for protein molecular dynamicists, applied to hydrogen bonds in the active site of elastase. J. Mol. Graphics*, **7**, 240–242, 223–224.

Takenaka, A. & Sasada, Y. (1980). *Computer manipulation of crystal and molecular models. J. Crystallogr. Soc. Jpn*, **22**, 214–225.

Thomas, D. J. (1993). *Toward more reliable printed stereo. J. Mol. Graphics*, **11**, 15–22.

Tsernoglou, D., Petsko, G. A., McQueen, J. E. & Hermans, J. (1977). *Molecular graphics: application to the structure determination of a snake venom neurotoxin. Science*, **197**, 1378–1381.

Vedani, A. & Meyer, E. F. (1984). *Structure–activity relationships of sulfonamide drugs and human carbonic anhydrase C: modelling of inhibitor molecules into receptor site of the enzyme with an interactive computer graphics display. J. Pharm. Sci.* **73**, 352–358.

Walsh, G. R. (1975). *Methods of Optimization.* London: John Wiley.

Warme, P. K., Go, N. & Scheraga, H. A. (1972). *Refinement of X-ray data of proteins. 1. Adjustment of atomic coordinates to conform to a specified geometry. J. Comput. Phys.* **9**, 303–317.

Watkin, D. J., Prout, C. K. & Pearce, L. J. (1996). *CAMERON*. Chemical Crystallography Laboratory, University of Oxford, England.

Williams, T. V. (1982). Thesis. University of North Carolina at Chapel Hill, NC, USA.

Willoughby, T. V., Morimoto, C. N., Sparks, R. A. & Meyer, E. F. (1974). *Mini-computer control of a stereo graphics display. J. Appl. Cryst.* **7**, 430–434.

Wipke, W. T. (1974). *Computer assisted three-dimensional synthetic analysis.* In *Computer Representation and Manipulation of Chemical Information*, edited by W. T. Wipke, S. R. Heller, R. J. Feldmann & E. Hyde, pp. 147–174. New York: John Wiley.

Wipke, W. T., Braun, H., Smith, G., Choplin, F. & Sieber, W. (1977). *SECS – simulation and evaluation of chemical synthesis: strategy and planning. ACS Symp. Ser.* **61**, 97–125.

Wipke, W. T. & Dyott, T. M. (1974). *Simulation and evaluation of chemical synthesis. Computer representation and manipulation of stereochemistry. J. Am. Chem. Soc.* **96**, 4825–4834.

Yamamoto, A. (1982). *Structure factor of modulated crystal structures. Acta Cryst.* A**38**, 87–92.

International Tables for Crystallography (2010). Vol. B, Chapter 3.4, pp. 449–457.

3.4. Accelerated convergence treatment of R^{-n} lattice sums

By D. E. Williams†

3.4.1. Introduction

The electrostatic energy of an ionic crystal is often represented by taking a pairwise sum between charge sites interacting *via* Coulomb's law (the $n = 1$ sum). The individual terms may be positive or negative, depending on whether the pair of sites have charges of the same or different signs. The Coulombic energy is very long-range, and it is well known that convergence of the Coulombic lattice-energy sum is extremely slow. For simple structure types Madelung constants have been calculated which represent the Coulombic energy in terms of the cubic lattice constant or a nearest-neighbour distance. Glasser & Zucker (1980) give tables of Madelung constants and review the subject giving references dating back to 1884. If the ionic crystal structure is not of a simple type usually no Madelung constant will be available and the Coulombic energy must be obtained for the specific crystal structure being considered. In carrying out this calculation, accelerated-convergence treatment of the Coulombic lattice sum is indispensable to achieve accuracy with a reasonable amount of computational effort. A model of a molecular crystal may include partial net atomic charges or other charge sites such as lone-pair electrons. The $(n = 1)$ sum also applies between these site charges.

The dispersion energy of ionic or molecular crystals may be represented by an $(n = 6)$ sum over atomic sites, with possible inclusion of $(n = 8, 10, \ldots)$ terms for higher accuracy. The dispersion-energy sum has somewhat better convergence properties than the Coulombic sum. Nevertheless, accelerated-convergence treatment of the dispersion sum is strongly recommended since its use can yield at least an order of magnitude improvement in accuracy for a given calculation effort. The repulsion energy between nonbonded atoms in a crystal may be represented by an exponential function of short range, or possibly by an $(n = 12)$ function of short range. The convergence of the repulsion energy is fast and no accelerated-convergence treatment is normally required.

3.4.2. Definition and behaviour of the direct-space sum

This pairwise sum is taken between atoms (or sites) in the reference unit cell and all other atoms (or sites) in the crystal, excluding the self terms. Thus, the second atom (or site) is taken to range over the entire crystal, with elimination of self-energy terms. If V_n represents an energy, each atom is assigned one half of the pair energy. Therefore, the energy per unit cell is

$$V_n = (1/2) \sum_{j}^{\text{one cell}} \sum_{k}^{\prime\ \text{all cells}} Q_{jk} R_{jk}^{-n},$$

where Q_{jk} is a given coefficient, R_{jk} is an interatomic distance, and the prime on the second sum indicates that self terms are omitted. In the case of the Coulombic sum, $n = 1$ and $Q_{jk} = q_j q_k$ is the product of the site charges.

Table 3.4.2.1 gives an example of the convergence behaviour of the untreated $(n = 1)$ Coulombic sum for sodium chloride. Even at the rather large summation limit of 20 Å the Coulombic lattice sum has not converged and is incorrect by about 8%. The 20 Å sum included 832 molecules and 2494 individual distances. At various smaller summation limits the truncation error fluctuates wildly and can be either positive or negative. Note that the results shown in the table always refer to summation over whole molecules, that is, over neutral charge units.

If the Coulombic summation is not carried out over neutral charge units the truncation error is even larger. These considerations support the conclusion that accelerated-convergence treatment of the Coulombic lattice sum should be regarded as mandatory. Table 3.4.2.2 gives an example of the convergence behaviour of the untreated $(n = 6)$ dispersion sum for benzene. In obtaining this sum it is not necessary to consider whole molecules as in the Coulombic case. The exclusion of atoms (or sites) in the portions of molecules outside the summation limit greatly reduces the number of terms to be considered. At the summation limit of 20 Å, 439 benzene molecules and 22 049 individual distances are considered; the dispersion-sum truncation error is 0.4%. Thus, if sufficient computer time is available it may be possible to obtain a moderately accurate dispersion sum without the use of accelerated convergence. However, as shown below, the use of accelerated convergence will greatly speed up the calculation, and is in practice necessary if higher accuracy is required.

3.4.3. Preliminary description of the method

Ewald (1921) developed a method which modified the mathematical representation of the Coulombic lattice sum to improve the rate of convergence. This method was based on partially transforming the lattice sum into reciprocal space. Bertaut (1952) presented another method for derivation of the Ewald result which used the concept of the crystallographic structure factor. His formula extended the Ewald treatment to a composite lattice with more than one atom per lattice point. Nijboer & DeWette (1957) developed a general Fourier transform method for the evaluation of R^{-n} sums in simple lattices. Williams (1971) extended this treatment to a composite lattice and gave general formulae for the R^{-n} sums for any crystal. A review article, on which this chapter is based, appeared later (Williams, 1989*a*,*b*).

Consider a function, $W(R)$, which is unity at $R = 0$ and smoothly declines to zero as R approaches infinity. If each term of the lattice sum is multiplied by $W(R)$, the rate of convergence is increased. However, the rate of convergence of the remainder of

Table 3.4.2.1. *Untreated lattice-sum results for the Coulombic energy (n = 1) of sodium chloride (kJ mol^{-1}, Å); the lattice constant is taken as 5.628 Å*

Truncation limit	Number of molecules	Number of terms	Calculated energy
6.0	23	67	−696.933
8.0	59	175	−597.371
10.0	108	322	−915.152
12.0	201	601	−773.475
14.0	277	829	−796.248
16.0	426	1276	−826.502
18.0	587	1759	−658.995
20.0	832	2494	−794.619
Converged value			−862.825

† Deceased. Questions related to this chapter may be addressed to Dr Bill Smith, Molecular Simulation Group, Computational Science and Engineering Department, CCLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, United Kingdom.

Copyright © 2010 International Union of Crystallography

Table 3.4.2.2. *Untreated lattice-sum results for the dispersion energy (n = 6) of crystalline benzene (kJ mol^{-1}, Å)*

Truncation limit	Number of molecules	Number of terms	Calculated energy
6.0	26	524	−69.227
8.0	51	1313	−76.007
10.0	77	2631	−78.179
12.0	126	4718	−79.241
14.0	177	7531	−79.726
16.0	265	11274	−80.013
18.0	344	15904	−80.178
20.0	439	22049	−80.295
Converged value			−80.589

the original sum, which contains the difference terms, is not increased.

$$V_n = (1/2) \sum_{j}^{\text{one cell}} \sum_{k}^{\prime\ \text{all cells}} Q_{jk} R_{jk}^{-n} W(R) + (1/2) \sum_{j}^{\text{one cell}} \sum_{k}^{\prime\ \text{all cells}} Q_{jk} R_{jk}^{-n} [1 - W(R)].$$

In the accelerated-convergence method the difference terms are expressed as an integral of the product of two functions. According to Parseval's theorem (described below) this integral is equal to an integral of the product of the two Fourier transforms of the functions. Finally, the integral over the Fourier transforms of the functions is converted to a sum in reciprocal (or Fourier-transform) space. The choice of the convergence function $W(R)$ is not unique; an obvious requirement is that the relevant Fourier transforms must exist and have correct limiting behaviour. Nijboer and DeWette suggested using the incomplete gamma function for $W(R)$. More recently, Fortuin (1977) showed that this choice of convergence function leads to optimal convergence of the sums in both direct and reciprocal space:

$$W(R) = \Gamma(n/2, \pi w^2 R^2)/\Gamma(n/2),$$

where $\Gamma(n/2)$ and $\Gamma(n/2, \pi w^2 R^2)$ are the gamma function and the incomplete gamma function, respectively:

$$\Gamma(n/2, \pi w^2 R^2) = \int_{\pi w^2 R^2}^{\infty} t^{(n/2)-1} \exp(-t)\, \mathrm{d}t$$

and

$$\Gamma(n/2) = \Gamma(n/2, 0).$$

The complement of the incomplete gamma function is

$$\gamma(n/2, \pi w^2 R^2) = \Gamma(n/2) - \Gamma(n/2, \pi w^2 R^2).$$

3.4.4. Preliminary derivation to obtain a formula which accelerates the convergence of an R^{-n} sum over lattice points X(d)

The three-dimensional direct-space crystal lattice is specified by the origin vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. A general vector in direct space is defined as

$$\mathbf{X}(\mathbf{x}) = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3,$$

where x_1, x_2, x_3 are the fractional cell coordinates of $\mathbf{X}$. A lattice vector in direct space is defined as

$$\mathbf{X}(\mathbf{d}) = d_1\mathbf{a}_1 + d_2\mathbf{a}_2 + d_3\mathbf{a}_3,$$

where d_1, d_2, d_3 are integers (specifying particular values of x_1, x_2, x_3) designating a lattice point. V_d is the direct-cell volume which is equal to $\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3$. A general point in the direct lattice is $\mathbf{X}(\mathbf{x})$; the contents of the lattice are by definition identical as the components of $\mathbf{x}$ are increased or decreased by integer amounts.

The reciprocal-lattice vectors are defined by the relations

$$\mathbf{a}_j \cdot \mathbf{b}_k = 1 \qquad j = k$$
$$= 0 \qquad j \neq k.$$

A general vector in reciprocal space $\mathbf{H}(\mathbf{r})$ is defined as

$$\mathbf{H}(\mathbf{r}) = r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + r_3\mathbf{b}_3.$$

A reciprocal-lattice vector $\mathbf{H}(\mathbf{h})$ is defined by the integer triplet h_1, h_2, h_3 (specifying particular values of r_1, r_2, r_3) so that

$$\mathbf{H}(\mathbf{h}) = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3.$$

In other sections of this volume a shortened notation $\mathbf{h}$ is used for the reciprocal-lattice vector. In this section the symbol $\mathbf{H}(\mathbf{h})$ is used to indicate that it is a particular value of $\mathbf{H}(\mathbf{r})$.

The three-dimensional Fourier transform $g(\mathbf{t})$ of a function $f(\mathbf{x})$ is defined by

$$g(\mathbf{t}) = FT_3[f(\mathbf{x})] = \int f(\mathbf{x}) \exp(2\pi i \mathbf{x} \cdot \mathbf{t})\, \mathrm{d}\mathbf{x}.$$

The Fourier transform of the set of points defining the direct lattice is the set of points defining the reciprocal lattice, scaled by the direct-cell volume. It is useful for our purpose to express the lattice transform in terms of the Dirac delta function $\delta(x - x_\mathrm{o})$ which is defined so that for any function $f(\mathbf{x})$

$$f(\mathbf{x}_\mathrm{o}) = \int \delta(\mathbf{x} - \mathbf{x}_\mathrm{o}) f(\mathbf{x})\, \mathrm{d}\mathbf{x}.$$

We then write

$$FT_3\{\sum_{\mathbf{d}} \delta[\mathbf{X}(\mathbf{x}) - \mathbf{X}(\mathbf{d})]\} = V_\mathrm{d}^{-1} \sum_{\mathbf{h}} \delta[\mathbf{H}(\mathbf{r}) - \mathbf{H}(\mathbf{h})].$$

First consider the lattice sum over the direct-lattice points $\mathbf{X}(\mathbf{d})$, relative to a particular point $\mathbf{X}(\mathbf{x}) = \mathbf{R}$, with omission of the origin lattice point.

$$S'(n, \mathbf{R}) = \sum_{\mathbf{d} \neq 0} |\mathbf{X}(\mathbf{d}) - \mathbf{R}|^{-n}.$$

The special case with $\mathbf{R} = 0$ will also be needed:

$$S'(n, 0) = \sum_{\mathbf{d} \neq 0} |\mathbf{X}(\mathbf{d})|^{-n}.$$

Now define a sum of Dirac delta functions

$$f'[\mathbf{X}(\mathbf{d})] = \sum_{d\neq 0}\delta[\mathbf{X}(\mathbf{x}) - \mathbf{X}(\mathbf{d})].$$

Then S' can be represented as an integral

$$S'(n, \mathbf{R}) = \int f'[\mathbf{X}(\mathbf{d})]|\mathbf{X} - \mathbf{R}|^{-n}\,\mathrm{d}\mathbf{X},$$

in which a term is contributed to S' whenever the direct-space vector $\mathbf{X}$ coincides with the lattice vector $\mathbf{X}(\mathbf{d})$, except for $\mathbf{d} = 0$. Now apply the convergence function to S':

$$\begin{aligned} S'(n, \mathbf{R}) = {} & [\Gamma(n/2)]^{-1}\int f'[\mathbf{X}(\mathbf{d})]|\mathbf{X} - \mathbf{R}|^{-n} \\ & \times \Gamma(n/2, \pi w^2|\mathbf{X} - \mathbf{R}|^2)\,\mathrm{d}\mathbf{X} \\ & + [\Gamma(n/2)]^{-1}\int f'[\mathbf{X}(\mathbf{d})]|\mathbf{X} - \mathbf{R}|^{-n} \\ & \times \gamma(n/2, \pi w^2|\mathbf{X} - \mathbf{R}|^2)\,\mathrm{d}\mathbf{X}. \end{aligned}$$

The first integral is shown here only for the purpose of giving a consistent representation of S'; in fact, the first integral will be reconverted back into a sum and evaluated in direct space. The second integral will be transformed to reciprocal space using Parseval's theorem [see, for example, Arfken (1970)], which states that

$$\int f(\mathbf{X})g^*(\mathbf{X})\,\mathrm{d}\mathbf{X} = \int FT_3[f(\mathbf{X})]FT_3[g^*(\mathbf{X})]\,\mathrm{d}\mathbf{H}.$$

Considering only the second integral in the formula for S' and explicitly introducing the $\mathbf{d} = 0$ term we have

$$\begin{aligned} & [\Gamma(n/2)]^{-1}\int f[\mathbf{X}(\mathbf{d})]|\mathbf{X}(\mathbf{d}) - \mathbf{R}|^{-n}\gamma(n/2, \pi w^2|\mathbf{X} - \mathbf{R}|^2)\,\mathrm{d}\mathbf{X} \\ & - [\Gamma(n/2)]^{-1}\int \delta(\mathbf{X})|\mathbf{R}|^{-n}\gamma(n/2, \pi w^2|\mathbf{R}|^2)\,\mathrm{d}\mathbf{X}, \end{aligned}$$

where the unprimed f includes the $\mathbf{h} = 0$ term which was earlier omitted from f':

$$f(\mathbf{X}) = \sum_{\mathbf{d}}\delta[\mathbf{X}(\mathbf{x}) - \mathbf{X}(\mathbf{d})].$$

Using Parseval's theorem, and evaluating the origin term, we have

$$\begin{aligned} & [\Gamma(n/2)]^{-1}\int FT_3\{f[\mathbf{X}(\mathbf{d})]\}FT_3 \\ & \quad \times [|\mathbf{X}(\mathbf{d}) - \mathbf{R}|^{-n}\gamma(n/2, \pi w^2|\mathbf{X} - \mathbf{R}|^2)]\,\mathrm{d}\mathbf{H} \\ & \quad - [\Gamma(n/2)]^{-1}|\mathbf{R}|^{-n}\gamma(n/2, \pi w^2|\mathbf{R}|^2). \end{aligned}$$

The Fourier transform of the complement of the incomplete gamma function divided by $|\mathbf{X}|^n$ is (Nijboer & DeWette, 1957)

$$\begin{aligned} & FT_3[\gamma(n/2, \pi w^2|\mathbf{X}|^2)|\mathbf{X}|^{-n}] \\ & \quad = \pi^{n-(3/2)}|\mathbf{H}|^{n-3}\Gamma[(-n/2) + (3/2), \pi w^{-2}|\mathbf{H}|^2]. \end{aligned}$$

If there is a change of origin and the point $(\mathbf{X} - \mathbf{R})$ is used instead of $\mathbf{X}$ the transform is

$$\begin{aligned} & FT_3[\gamma(n/2, \pi w^2|\mathbf{X} - \mathbf{R}|^2)|\mathbf{X} - \mathbf{R}|^{-n}] \\ & \quad = \pi^{n-(3/2)}|\mathbf{H}|^{n-3}\Gamma[(-n/2) \\ & \qquad + (3/2), \pi w^{-2}|\mathbf{H}|^2]\exp(2\pi i\mathbf{H}\cdot\mathbf{R}). \end{aligned}$$

Evaluation of the two Fourier transforms in the first term gives

$$\begin{aligned} & [\Gamma(n/2)]^{-1}\int V_{\mathrm{d}}^{-1}\sum_{\mathbf{h}}\delta[\mathbf{H}(\mathbf{h}) - \mathbf{H}]\pi^{n-(3/2)}|\mathbf{H}|^{n-3} \\ & \quad \times \Gamma[(-n/2) + (3/2), \pi w^{-2}|\mathbf{H}|^2]\exp(2\pi i\mathbf{H}\cdot\mathbf{R})\,\mathrm{d}\mathbf{H}. \end{aligned}$$

Because of the presence of the Dirac delta function in each integral, we can convert the integrals with $\mathbf{h}$ unequal to zero into a sum

$$\begin{aligned} & [\Gamma(n/2)]^{-1}V_{\mathrm{d}}^{-1}\pi^{n-(3/2)}\sum_{h\neq 0}|\mathbf{H}(\mathbf{h})|^{n-3} \\ & \quad \times \Gamma[(-n/2) + (3/2), \pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]\exp[2\pi i\mathbf{H}(\mathbf{h})\cdot\mathbf{R}]. \end{aligned}$$

The $\mathbf{h} = 0$ term needs to be evaluated in the limit. Clearly, the complex exponential goes to unity. If n is greater than 3 the limit of the indeterminate form infinity/infinity is needed:

$$\begin{aligned} & \lim_{|\mathbf{H}|\to 0}\frac{\Gamma[(-n/2) + (3/2), \pi w^{-2}|\mathbf{H}|^2]}{|\mathbf{H}|^{3-n}} \\ & \quad = \lim_{|\mathbf{H}|\to 0}\frac{\int_{\pi w^{-2}|\mathbf{H}|^2}^{\infty} t^{(-n/2)+(1/2)}\exp(-t)\,\mathrm{d}t}{|\mathbf{H}|^{3-n}}. \end{aligned}$$

The limit can be found by L'Hospital's rule [see, for example, Widder (1961)] which states that if $f(x)$ and $g(x)$ both approach infinity as x approaches a constant, c, and the limit of the ratio of the first derivatives $f'(x)$ and $g'(x)$ exists, that limit is also true for the limit of the ratio of the functions:

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}.$$

To differentiate the definite integral function, Leibnitz's formula may be used [see, for example, Arfken (1970)]. This formula states that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\int_{g(x)}^{h(x)} f(t, x)\,\mathrm{d}t = {} & \int_{g(x)}^{h(x)}\frac{\mathrm{d}f(t, x)}{\mathrm{d}x}\,\mathrm{d}t \\ & + f[h(x)]\frac{\mathrm{d}h(x)}{\mathrm{d}x} - f[g(x)]\frac{\mathrm{d}g(x)}{\mathrm{d}x}. \end{aligned}$$

In our case, x becomes $|\mathbf{H}|$; f becomes $t^{(-n/2)+(1/2)}\exp(-t)$ which is independent of $|\mathbf{H}|$; g becomes $\pi w^{-2}|\mathbf{H}|^2$; and h is infinite. Thus only the last term of Leibnitz's formula is nonzero and we obtain for the ratio of the first derivatives

$$\begin{aligned} & \lim_{|\mathbf{H}|\to 0}\frac{-(\pi w^{-2}|\mathbf{H}|^2)^{(-n/2)+(1/2)}\exp(-\pi w^2|\mathbf{H}|^2)2\pi w^{-2}|\mathbf{H}|}{(3-n)|\mathbf{H}|^{2-n}} \\ & \quad = \pi^{(-n/2)+(3/2)}w^{n-3}[2/(n-3)], \end{aligned}$$

so that the limiting value for the $\mathbf{h} = 0$ term for n greater than 3 is

$$+[\Gamma(n/2)]^{-1}V_{\rm d}^{-1}\pi^{n/2}w^{n-3}[2/(n-3)].$$

The final result for S' is

$$\begin{aligned}S'(n,\mathbf{R}) &= [\Gamma(n/2)]^{-1}\sum_{\mathbf{d}\neq 0}|\mathbf{X}(\mathbf{d})-\mathbf{R}|^{-n}\\ &\times\Gamma(n/2,\pi w^2|\mathbf{X}(\mathbf{d})-\mathbf{R}|^2)\\ &-[\Gamma(n/2)]^{-1}|\mathbf{R}|^{-n}\gamma(n/2,\pi w^2|\mathbf{R}|^2)\\ &+[\Gamma(n/2)]^{-1}V_{\rm d}^{-1}\pi^{n-(3/2)}\sum_{\mathbf{h}\neq 0}|\mathbf{H}(\mathbf{h})|^{n-3}\\ &\times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]\\ &\times\exp[2\pi i\mathbf{H}(\mathbf{h})\cdot\mathbf{R}]\\ &+[\Gamma(n/2)]^{-1}V_{\rm d}^{-1}\pi^{n/2}w^{n-3}2(n-3)^{-1}.\end{aligned}$$

The significance of the terms is as follows. The first term represents the convergence-accelerated direct sum, which does not include the origin term; the next term, also in direct space, corrects for the remainder resulting from the subtraction of the origin term; the third term comes from Parseval's theorem and is a sum over the nonzero **h** reciprocal-lattice points; and the last term is the reciprocal-lattice $\mathbf{h}=0$ term.

If $\mathbf{R}=0$ the second term becomes an indeterminate form 0/0. The limit can be found with use of L'Hospital's rule again, this time for the 0/0 form. We need the limit of $f(x)/g(x)$, where $f(R)=\gamma(n/2,\pi w^2R^2)$ and $g(R)=R^n$. To differentiate the incomplete gamma function, we can again use Leibnitz's formula. In this case only the second term of the formula is nonzero and we obtain for the ratio of the first derivatives

$$\frac{2\pi^{n/2}w^n|\mathbf{R}|^{n-1}\exp(-\pi w^2|\mathbf{R}|^2)}{n|\mathbf{R}|^{n-1}},$$

so that the limiting value for this term as $|\mathbf{R}|$ approaches zero is

$$-[\Gamma(n/2)]^{-1}2\pi^{n/2}w^n n^{-1}.$$

Therefore, the value of the sum when $\mathbf{R}=0$ is

$$\begin{aligned}S'(n,0) &= [\Gamma(n/2)]^{-1}\sum_{\mathbf{d}\neq 0}|\mathbf{X}(\mathbf{d})|^{-n}\Gamma(n/2,\pi w^2|\mathbf{X}(\mathbf{d})|^2)\\ &-[\Gamma(n/2)]^{-1}2\pi^{n/2}w^n n^{-1}\\ &+[\Gamma(n/2)]^{-1}V_{\rm d}^{-1}\pi^{n-(3/2)}\sum_{\mathbf{h}\neq 0}|\mathbf{H}(\mathbf{h})|^{n-3}\\ &\times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]\\ &+[\Gamma(n/2)]^{-1}V_{\rm d}^{-1}\pi^{n/2}w^{n-3}2(n-3)^{-1}.\end{aligned}$$

3.4.5. Extension of the method to a composite lattice

Define a general lattice sum over direct-space points $\mathbf{R}_j$ which interact with pairwise coefficients Q_{jk}, where $Q_{jk}=Q_{kj}$:

$$V(n,\mathbf{R}_j)=(1/2)\sum_j{\sum_k}'\,Q_{jk}\sum_{\mathbf{d}}|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^{-n},$$

where the prime indicates that when $\mathbf{d}=0$ the self-terms with $j=k$ are omitted. For convenience the terms may be divided into three groups: the first group of terms has $\mathbf{d}=0$, where j is unequal to k; the second group has **d** not zero and j not equal to k; and the third group had **d** not zero and $j=k$. (A possible fourth group with $\mathbf{d}=0$ and $j=k$ is omitted, as defined.)

$$\begin{aligned}V(n,\mathbf{R}_j) &= (1/2)\sum_{j\neq k}Q_{jk}|\mathbf{R}_k-\mathbf{R}_j|^{-n}\\ &+(1/2)\sum_{j\neq k}Q_{jk}S'(n,|\mathbf{R}_j-\mathbf{R}_k|)+(1/2)\sum_j Q_{jj}S'(n,0).\end{aligned}$$

By expanding this expression we obtain

$$\begin{aligned}V(n,\mathbf{R}_j) &= (1/2)\sum_{j\neq k}Q_{jk}|\mathbf{R}_k-\mathbf{R}_j|^{-n} &(1)\\ &+\Big\{[1/2\Gamma(n/2)]\sum_{j\neq k}Q_{jk}\sum_{\mathbf{d}\neq 0}|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^{-n}\\ &\times\Gamma(n/2,\pi w^2|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^2)\Big\} &(2)\\ &-\Big\{[1/2\Gamma(n/2)]\sum_{j\neq k}Q_{jk}|\mathbf{R}_k-\mathbf{R}_j|^{-n}\\ &\times\gamma(n/2,\pi w^2|\mathbf{R}_k-\mathbf{R}_j|^2)\Big\} &(3)\\ &+\Big\{[1/2\Gamma(n/2)]V_{\rm d}^{-1}\pi^{n-(3/2)}\sum_{j\neq k}Q_{jk}\sum_{\mathbf{h}\neq 0}|\mathbf{H}(\mathbf{h})|^{n-3}\\ &\times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]\\ &\times\exp[2\pi i\mathbf{H}(\mathbf{h})\cdot(\mathbf{R}_k-\mathbf{R}_j)]\Big\} &(4)\\ &+[1/2\Gamma(n/2)]V_{\rm d}^{-1}\pi^{n/2}w^{n-3}2(n-3)^{-1}\sum_{j\neq k}Q_{jk} &(5)\\ &+\Big\{[1/2\Gamma(n/2)]\sum_j Q_{jj}\sum_{\mathbf{d}\neq 0}|\mathbf{X}(\mathbf{d})|^{-n}\\ &\times\Gamma(n/2,\pi w^2|\mathbf{X}(\mathbf{d})|^2)\Big\} &(6)\\ &-[1/\Gamma(n/2)]\pi^{n/2}w^n n^{-1}\sum_j Q_{jj} &(7)\\ &+\Big\{[1/2\Gamma(n/2)]V_{\rm d}^{-1}\pi^{n-(3/2)}\sum_j Q_{jj}\sum_{\mathbf{h}\neq 0}|\mathbf{H}(\mathbf{h})|^{n-3}\\ &\times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]\Big\} &(8)\\ &+[1/2\Gamma(n/2)]V_{\rm d}^{-1}\pi^{n/2}w^{n-3}2(n-3)^{-1}\sum_j Q_{jj}. &(9)\end{aligned}$$

This expression for V has nine terms, which are numbered on the right-hand side. Term (3) can be expressed in terms of Γ rather than γ:

$$\begin{aligned}(3) &= -(1/2)\sum_{j\neq k}Q_{jk}|\mathbf{R}_k-\mathbf{R}_j|^{-n}\\ &+[1/\Gamma(n/2)]\sum_{j\neq k}Q_{jk}|\mathbf{R}_k-\mathbf{R}_j|^{-n}\Gamma(n/2,\pi w^2|\mathbf{R}_k-\mathbf{R}_j|^2).\end{aligned}$$

It is seen that cancellation occurs with term (1) so that

$$(1)+(3)=[1/\Gamma(n/2)]\sum_{j\neq k} Q_{jk}|\mathbf{R}_k-\mathbf{R}_j|^{-n} \times \Gamma(n/2,\pi w^2|\mathbf{R}_k-\mathbf{R}_j|^2),$$

which is the $\mathbf{d}=0$, j unequal to k portion of the treated direct-lattice sum. The $\mathbf{d}$ unequal to 0, j unequal to k portion corresponds to term (2) and the $\mathbf{d}$ unequal to 0, $j=k$ portion corresponds to term (6). The direct-lattice terms may be consolidated as

$$(1)+(2)+(3)+(6)=[1/2\Gamma(n/2)]\sum_j{\sum_k}' Q_{jk}\sum_{\mathbf{d}}|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^{-n} \times \Gamma[n/2,\pi w^2|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^2].$$

Now let us combine terms (4) and (8), carrying out the $\mathbf{h}$ summation first:

$$(4)+(8)=[1/2\Gamma(n/2)]V_{\mathrm{d}}^{-1}\pi^{n-(3/2)}\sum_{\mathbf{h}}|\mathbf{H}(\mathbf{h})|^{n-3} \times \Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2] \times \sum_j\sum_k Q_{jk}\exp[2\pi i\mathbf{H}(\mathbf{h})\cdot(\mathbf{R}_k-\mathbf{R}_j)].$$

Terms (5) and (9) may be combined:

$$(5)+(9)=[\Gamma(n/2)]^{-1}V_{\mathrm{d}}^{-1}\pi^{n/2}w^{n-3}(n-3)^{-1}\left(\sum_j Q_{ij}+\sum_{j\neq k} Q_{jk}\right).$$

The final formula is shown below. The significance of the four terms is: (1) the treated direct-lattice sum; (2) a correction for the difference resulting from the removal of the origin term in direct space; (3) the reciprocal-lattice sum, except $\mathbf{h}=0$; and (4) the $\mathbf{h}=0$ term of the reciprocal-lattice sum.

$$\begin{aligned}
&V(n,\mathbf{R}_j)\\
&\quad=[1/2\Gamma(n/2)]\sum_j{\sum_k}' Q_{jk}\sum_{\mathbf{d}}|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^{-n}\\
&\qquad\times\Gamma(n/2,\pi w^2|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^2) \qquad (1)\\
&\qquad-[1/\Gamma(n/2)]\pi^{n/2}w^n n^{-1}\sum_j Q_{jj} \qquad (2)\\
&\qquad+[1/2\Gamma(n/2)]V_{\mathrm{d}}^{-1}\pi^{n-(3/2)}\sum_{\mathbf{h}}|\mathbf{H}(\mathbf{h})|^{n-3}\\
&\qquad\times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]\\
&\qquad\times\sum_j\sum_k Q_{jk}\exp[2\pi i\mathbf{H}(\mathbf{h})\cdot(\mathbf{R}_k-\mathbf{R}_j)] \qquad (3)\\
&\qquad+[\Gamma(n/2)]^{-1}V_{\mathrm{d}}^{-1}\pi^{n/2}w^{n-3}(n-3)^{-1}\left(\sum_j\sum_k Q_{jk}\right). \qquad (4)
\end{aligned}$$

3.4.6. The case of $n=1$ (Coulombic lattice energy)

As taken above, the limit of the reciprocal-lattice $\mathbf{h}=0$ term of $S'(n,\mathbf{R})$ or $S'(n,0)$ existed only if n was greater than 3. The corresponding contributions to $V(n,\mathbf{R}_j)$ were terms (5) and (9) of Section 3.4.5. To extend the method to $n=1$ we will show in this section that these $\mathbf{h}=0$ terms vanish if conditions of unit-cell neutrality and zero dipole moment are satisfied.

The integral representation of the term (5) is

$$[1/2\Gamma(n/2)]V_{\mathrm{d}}^{-1}\pi^{n-(3/2)}\sum_{j\neq k} Q_{jk}\int\delta(0)|\mathbf{H}|^{n-3} \times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}|^2] \times\exp[2\pi i\mathbf{H}\cdot(\mathbf{R}_k-\mathbf{R}_j)]\,\mathrm{d}\mathbf{H}$$

and for term (9) is

$$[1/2\Gamma(n/2)]V_{\mathrm{d}}^{-1}\pi^{n-(3/2)}\sum_j Q_{jj}\int\delta(0)|\mathbf{H}|^{n-3} \times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}|^2]\,\mathrm{d}\mathbf{H}.$$

Combining these two sums of integrals into one integral sum gives

$$[1/2\Gamma(n/2)]V_{\mathrm{d}}^{-1}\pi^{n-(3/2)}\int\delta(0)|\mathbf{H}|^{n-3} \times\Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}|^2]\sum_j{\sum_k} Q_{jk} \times\exp[2\pi i\mathbf{H}\cdot(\mathbf{R}_k-\mathbf{R}_j)]\,\mathrm{d}\mathbf{H}.$$

For $n=1$, suppose q_j are net atomic charges so that the geometric combining law holds for $Q_{jk}=q_jq_k$. Then the double sum over j and k can be factored so that the limit that needs to be considered is

$$\lim_{|\mathbf{H}|\to 0}\frac{\left[\sum_k q_k\exp(2\pi i\mathbf{H}\cdot\mathbf{R}_k)\right]\left[\sum_j q_j\exp(-2\pi i\mathbf{H}\cdot\mathbf{R}_j)\right]}{|\mathbf{H}|^2}.$$

If the unit cell does not have a net charge, the sum over the q's goes to zero in the limit and this is a 0/0 indeterminate form. Let $|\mathbf{H}|$ approach zero along the polar axis so that $\mathbf{H}\cdot\mathbf{R}_k=H_3R_{3k}$, where subscript 3 indicates components along the polar axis. To find the limit with L'Hospital's rule the numerator and denominator are differentiated twice with respect to H_3. Represent the numerator of the limit by the product (uv) and note that

$$\frac{\mathrm{d}^2(uv)}{\mathrm{d}x^2}=u\frac{\mathrm{d}^2v}{\mathrm{d}x^2}+v\frac{\mathrm{d}^2u}{\mathrm{d}x^2}+2\frac{\mathrm{d}u}{\mathrm{d}x}\frac{\mathrm{d}v}{\mathrm{d}x}.$$

It is seen that in addition to cell neutrality the product of the first derivatives of the sums must exist. These sums are

$$\left[2\pi i\sum_k q_kR_{3k}\exp(2\pi iH_3R_{3k})\right]$$

and

$$\left[-2\pi i\sum_j q_jR_{3j}\exp(-2\pi iH_3R_{3j})\right],$$

which vanish if the unit cell has no dipole moment in the polar direction, that is, if $\sum q_jR_{3j}=0$. Since the second derivative of the denominator is a constant, the desired limit is zero under the specified conditions. Now the polar direction can be chosen arbitrarily, so the unit cell must not have a dipole moment in any direction for the limit of the numerator to be zero. Thus we have the formula for the Coulombic lattice sum

$$\begin{aligned} V(1,\mathbf{R}_j) = {} & [1/2\Gamma(1/2)]\sum_j{\sum_k}' Q_{jk}\sum_{\mathbf{d}}|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^{-1} \\ & \times \Gamma(1/2,\pi w^2|\mathbf{R}_k+\mathbf{X}(\mathbf{d})-\mathbf{R}_j|^2) \\ & + [1/2\Gamma(1/2)V_{\rm d}^{-1}\pi^{-1/2}\sum_{\mathbf{h}}|\mathbf{H}(\mathbf{h})|^{-2} \\ & \times \Gamma(1/2,\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2)\sum_j\sum_k Q_{jk} \\ & \times \exp[2\pi i\mathbf{H}(\mathbf{h})\cdot(\mathbf{R}_k-\mathbf{R}_j)] \\ & - [1/\Gamma(1/2)]\pi^{1/2}w\sum_j q_j^2, \end{aligned}$$

which holds on conditions that the unit cell be electrically neutral and have no dipole moment. If the unit cell has a dipole moment, the limiting value discussed above depends on the direction of **H**. For methods of obtaining the Coulombic lattice sum where the unit cell does have a dipole moment, the reader is referred to the literature (DeWette & Schacher, 1964; Cummins *et al.*, 1976; Bertaut, 1978; Massidda, 1978).

3.4.7. The cases of $n = 2$ and $n = 3$

If $n = 2$ the denominator considered for the limit in the preceding section is linear in $|\mathbf{H}|$ so that only one differentiation is needed to obtain the limit by L'Hospital's method. Since a term of the type $\sum q_j \exp(2\pi i\mathbf{H}\cdot\mathbf{R}_j)$ is always a factor, the requirement that the unit cell have no dipole moment can be relaxed. For $n = 2$ the zero-charge condition is still required: $\sum q_j = 0$. When $n = 3$ the expression becomes determinate and no differentiation is required to obtain a limit. In addition, factoring the Q_{jk} sums into q_j sums is not necessary so that the only remaining requirement for this term to be zero is $\sum\sum Q_{jk} = 0$, which is a further relaxation beyond the requirement of cell neutrality.

3.4.8. Derivation of the accelerated convergence formula *via* the Patterson function

The structure factor with generalized coefficients q_j is defined by

$$F[\mathbf{H}(\mathbf{h})] = \sum_j q_j \exp[2\pi i\mathbf{H}(\mathbf{h})\cdot\mathbf{R}_j].$$

The corresponding Patterson function is defined by

$$P(\mathbf{X}) = V_{\rm d}^{-1}\sum_{\mathbf{h}}|F[\mathbf{H}(\mathbf{h})]|^2\exp[2\pi i\mathbf{H}(\mathbf{h})\cdot\mathbf{X}].$$

The physical interpretation of the Patterson function is that it is nonzero only at the intersite vector points $\mathbf{R}_k+\mathbf{X}(\mathbf{h})-\mathbf{R}_j$. If the origin point is removed, the lattice sum may be expressed as an integral over the Patterson function. This origin point in the Patterson function corresponds to intersite vectors with $j = k$ and $\mathbf{H}(\mathbf{h}) = 0$:

$$S_n = (1/2V_{\rm d})\int|\mathbf{X}|^{-n}[P(\mathbf{X})-P(\mathbf{X})\delta(\mathbf{X})]\,{\rm d}\mathbf{X}.$$

Using the incomplete gamma function as a convergence function, this formula expands into two integrals

$$\begin{aligned} S_n = {} & [1/2V_{\rm d}\Gamma(n/2)]\int|\mathbf{X}|^{-n} \\ & \times [P(\mathbf{X})-P(\mathbf{X})\delta(\mathbf{X})]\Gamma(n/2,\pi w^2|\mathbf{X}|^2)\,{\rm d}\mathbf{X} \\ & + [1/2V_{\rm d}\Gamma(n/2)]\int|\mathbf{X}|^{-n} \\ & \times [P(\mathbf{X})-P(\mathbf{X})\delta(\mathbf{X})]\gamma(n/2,\pi w^2|\mathbf{X}|^2)\,{\rm d}\mathbf{X}. \end{aligned}$$

The first integral is shown only for a consistent representation; actually it will be reconverted to a sum and evaluated in direct space. The first part of the second integral will be evaluated with Parseval's theorem and the second part in the limit as $|\mathbf{X}|$ approaches zero:

$$\begin{aligned} & [1/2V_{\rm d}\Gamma(n/2)]\int FT_3[P(\mathbf{X})] \\ & \quad \times FT_3[|\mathbf{X}|^{-n}\gamma(n/2,\pi w^2|\mathbf{X}|^2)]\,{\rm d}\mathbf{H} \\ & \quad - \lim_{\mathbf{X}\to 0}[1/2V_{\rm d}\Gamma(n/2)][P(0)|\mathbf{X}|^{-n}\gamma(n/2,\pi w^2|\mathbf{X}|^2)]. \end{aligned}$$

The first Fourier transform (of the Patterson function) is the set of amplitudes of the structure factors and the second Fourier transform has already been discussed above; the method for obtaining the limit (for n equal to or greater than 1) was also discussed above. The result obtained is

$$\begin{aligned} & [1/2V_{\rm d}\Gamma(n/2)]\pi^{n-(3/2)}\int|F[\mathbf{H}(\mathbf{h})]|^2|\mathbf{H}|^{n-3} \\ & \quad \times \Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}|^2]\,{\rm d}\mathbf{H} \\ & \quad - [1/2V_{\rm d}\Gamma(n/2)]|F(0)|^2 2\pi^{n/2}w^n n^{-1}. \end{aligned}$$

The integral can be converted into a sum, since $|F[\mathbf{H}(\mathbf{h})]|$ is nonzero only at the reciprocal-lattice points:

$$\begin{aligned} & [1/2V_{\rm d}\Gamma(n/2)]\pi^{n-(3/2)}\sum_{\mathbf{h}}|F[\mathbf{H}(\mathbf{h})]|^2|\mathbf{H}(\mathbf{h})|^{n-3} \\ & \quad \times \Gamma[(-n/2)+(3/2),\pi w^{-2}|\mathbf{H}(\mathbf{h})|^2]. \end{aligned}$$

The term with $\mathbf{H}(\mathbf{h}) = 0$ is evaluated in the limit, for n greater than 3, as

$$[\Gamma(n/2)]^{-1}V_{\rm d}^{-1}\pi^{n/2}w^{n-3}(n-3)^{-1}|F(0)|^2.$$

Since $|F(0)|^2 = \sum\sum q_j q_k$, this term is identical with the third term of $V(n,\mathbf{R}_j)$ as derived earlier. The case of $n = 1$ is handled in the same way as previously discussed, where the limit of this term is zero provided the unit cell has no net charge or dipole moment.

3.4.9. Evaluation of the incomplete gamma function

The incomplete gamma function may be expressed in terms of commonly available functions such as the exponential integral and the complement of the error function. The definition of the exponential integral is

$$E_1(x^2) = \int_{x^2}^{\infty} t^{-1}\exp(-t)\,{\rm d}t = \Gamma(0,x^2).$$

The definition of the complement of the error function is

$$\mathrm{erfc}(x) = \int_x^{\infty}\exp(-t^2)\,{\rm d}t = \pi^{-1/2}\Gamma(1/2,x^2).$$

Numerical approximations to these functions are given, for example, by Hastings (1955). The recursion formula for the incomplete gamma function (Davis, 1972)

$$\Gamma(n+1, x^2) = n\Gamma(n, x^2) + x^{2n}\exp(-x^2)$$

may be used to obtain working formulae starting from the special values of $\Gamma(0, x^2)$ and $\Gamma(1/2, x^2)$ which are defined above. Also we note that $\Gamma(1, x^2) = \exp(-x^2)$.

3.4.10. Summation over the asymmetric unit and elimination of intramolecular energy terms

Let us consider the case where the unit cell contains Z molecules which are related by Z symmetry operations, and it is desired to include only intermolecular distances in the summation. In the direct sum (1) the indices j and k will then run only over the asymmetric unit, and all terms with $\mathbf{d} = 0$ are eliminated. The calculated energy refers then to one molecule (or mole) rather than to one unit cell. The correction term (2) also refers to one molecule according to the range of j and k. Since the reciprocal-lattice sum refers to the entire unit cell, terms (3) and (4) need to be divided by Z to refer the energy to one molecule.

Both the direct and reciprocal sums must be corrected for the elimination of intramolecular terms. Using the convergence function $W(R)$, we have

$$\begin{aligned} V(n, \mathbf{R}_j) = {} & \sum_{\text{inter}} |\mathbf{R}|^{-n} W(R) + \sum_{\text{intra}} |\mathbf{R}|^{-n} W(R) \\ & + \sum_{\text{inter}} |\mathbf{R}|^{-n}[1 - W(R)] + \sum_{\text{intra}} |\mathbf{R}|^{-n}[1 - W(R)]. \end{aligned}$$

As mentioned above, the second summation term, which is the intramolecular term in direct space, is simply left out of the calculation. When using the accelerated-convergence method the third and fourth summation terms are always obtained, evaluated in reciprocal space. The undesired inclusion of the intramolecular term (fourth term above) in the reciprocal-space sum may be compensated for by explicit subtraction of this term from the sum.

3.4.11. Reference formulae for particular values of *n*

In this section let $a^2 = \pi w^2 |\mathbf{R}_k + \mathbf{X}(\mathbf{d}) - \mathbf{R}_j|^2$ and $b^2 = \pi w^{-2} |\mathbf{H}(\mathbf{h})|^2$. Let $T_0 = \sum Q_{jj} = \sum q_j^2$; $T_1 = \sum\sum Q_{jk} = T_0 + 2\sum\sum_{j>k} Q_{jk}$. If the geometric mean combining law holds, $T_1 = (\sum_j q_j)^2$; let

$$\begin{aligned} T_2(h) &= \sum_j \sum_k Q_{jk} \exp[2\pi i \mathbf{H}(\mathbf{h}) \cdot (\mathbf{R}_k - \mathbf{R}_j)] \\ &= T_0 + 2\sum_{j>k}\sum Q_{jk} \cos[2\pi \mathbf{H}(\mathbf{h}) \cdot (\mathbf{R}_k - \mathbf{R}_j)]. \end{aligned}$$

Then

$$T_2(\mathbf{h}) = |F(\mathbf{h})|^2 = \left| \sum_j q_j \exp[2\pi i \mathbf{H}(\mathbf{h}) \cdot \mathbf{R}_j] \right|^2 = A(\mathbf{h})^2 + B(\mathbf{h})^2,$$

where

$$A(\mathbf{h}) = \sum_j q_j \cos[2\pi \mathbf{H}(\mathbf{h}) \cdot \mathbf{R}_j]$$

and

$$B(\mathbf{h}) = \sum_j q_j \sin[2\pi \mathbf{H}(\mathbf{h}) \cdot \mathbf{R}_j].$$

The formulae below describe $V(n, \mathbf{R}_j)$ in terms of T_0, T_1 and T_2; the distance $|\mathbf{R}_k + \mathbf{X}(\mathbf{d}) - \mathbf{R}_j|$ is simply represented by R_{jkd}.

$$\begin{aligned} V(1, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k q_j\, q_k \sum_{\mathbf{d}} R_{jkd}^{-1} \operatorname{erfc}(a) \\ & + (1/2\pi V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})|^{-2} \exp(-b^2) - wT_0 \\ V(2, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-2} \exp(-a^2) \\ & + (\pi/2V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})|^{-1} \operatorname{erfc}(b) - (\pi/2) w^2 T_0 \\ V(3, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-3} [\operatorname{erfc}(a) + 2a\pi^{-1/2} \exp(-a^2)] \\ & + (\pi/V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) E_1(b^2) - (2\pi/3) w^3 T_0 \\ V(4, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-4} (1 + a^2) \exp(-a^2) \\ & + (\pi^{5/2}/V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})| \\ & \times [-\pi^{1/2} \operatorname{erfc}(b) + b^{-1} \exp(-b^2)] \\ & + (\pi^2/V_{\mathrm{d}}) w T_1 - (\pi^2/4) w^4 T_0 \\ V(5, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-5} \\ & \times [\operatorname{erfc}(a) + 2\pi^{-1/2} a (1 + 2a^2/3) \exp(-a^2)] \\ & + (2\pi^3/3V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})|^2 \\ & \times [b^{-2} \exp(-b^2) - E_1(b^2)] \\ & + (2\pi^2/3V_{\mathrm{d}}) w^2 T_1 - (4\pi^2/15) w^5 T_0 \\ V(6, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-6} [1 + a^2 + (a^4/2)] \exp(-a^2) \\ & + (\pi^{9/2}/3V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})|^3 \\ & \times [\pi^{1/2} \operatorname{erfc}(b) + [(1/2b^3) - (1/b)] \exp(-b^2)] \\ & + (\pi^3/6V_{\mathrm{d}}) w^3 T_1 - (\pi^3/12) w^6 T_0 \\ V(7, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-7} \\ & \times [\operatorname{erfc}(a) + 2\pi^{-1/2} a [1 + (2/3)a^2 + (4/15)a^4] \\ & \times \exp(-a^2)] \\ & + (2\pi^5/15V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})|^4 \\ & \times [(-b^{-2} + b^{-4}) \exp(-b^2) + E_1(b^2)] \\ & + (2\pi^3/15V_{\mathrm{d}}) w^4 T_1 - (8\pi^3/105) w^7 T_0 \\ V(8, \mathbf{R}_j) = {} & (1/2)\sum_j{}' \sum_k Q_{jk} \sum_{\mathbf{d}} R_{jkd}^{-8} \\ & \times [1 + a^2 + (a^4/2) + (a^6/6)] \exp(-a^2) \\ & + (2\pi^{13/2}/45V_{\mathrm{d}}) \sum_{\mathbf{h}\neq 0} T_2(\mathbf{h}) |\mathbf{H}(\mathbf{h})|^5 \\ & \times [-\pi^{1/2} \operatorname{erfc}(b) + [(1/b) - (1/2b^3) + (3/4b^5)] \\ & \times \exp(-b^2)] \\ & + (\pi^4/30V_{\mathrm{d}}) w^5 T_1 - (\pi^4/48) w^8 T_0 \end{aligned}$$

Table 3.4.12.1. *Accelerated-convergence results for the Coulombic sum (n = 1) of sodium chloride (kJ mol^{-1}, Å): the direct sum plus the constant term*

Limit	$w = 0.1$	$w = 0.15$	$w = 0.2$	$w = 0.3$	$w = 0.4$
6.0	−779.087	−838.145	−860.393	−924.275	−1125.372
8.0	−818.549	−860.194	−863.764	−924.282	−1125.372
10.0	−865.323	−862.818	−863.811	−924.282	
12.0	−861.183	−862.824	−863.811		
14.0	−862.717	−862.828			
16.0	−862.792	−862.828			
18.0	−862.810				
20.0	−862.825				

Table 3.4.12.2. *The reciprocal-lattice results (kJ mol^{-1}, Å) for the Coulombic sum (n = 1) of sodium chloride*

Limit	$w = 0.1$	$w = 0.15$	$w = 0.2$	$w = 0.3$	$w = 0.4$
0.0	−277.872	−416.806	−555.742	−833.613	−1111.483
0.4	0.000	0.003	0.986	61.451	261.042
0.5		0.003	0.986	61.451	261.042
0.6				61.457	262.542
0.7				61.457	262.542
0.8					262.547
0.9					262.547

$$
\begin{aligned}
\boldsymbol{V}(9,\mathbf{R}_j) &= (1/2)\sum_j{\sum_k}' Q_{jk}\sum_{\mathbf{d}} R_{jkd}^{-9} \\
&\quad\times[\operatorname{erfc}(a) + 2\pi^{-1/2}a[1 + (2/3)a^2 + (4/15)a^4 \\
&\quad + (8/105)a^6]\exp(-a^2)] \\
&\quad + (4\pi^7/315V_{\mathrm{d}})\sum_{\mathbf{h}\neq 0} T_2(\mathbf{h})|\mathbf{H}(\mathbf{h})|^6 \\
&\quad\times[(b^{-2} - b^{-4} + 2b^{-6})\exp(-b^2) - E_1(b^2)] \\
&\quad + (8\pi^4/315V_{\mathrm{d}})w^6T_1 - (16\pi^4/945)w^9T_0 \\
\boldsymbol{V}(10,\mathbf{R}_j) &= (1/2)\sum_j{\sum_k}' Q_{jk}\sum_{\mathbf{d}} R_{jkd}^{-10} \\
&\quad\times[1 + a^2 + (a^4/2) + (a^6/6) + (a^8/24)]\exp(-a^2) \\
&\quad + (\pi^{17/2}/315V_{\mathrm{d}})\sum_{\mathbf{h}\neq 0} T_2(\mathbf{h})|\mathbf{H}(\mathbf{h})|^7 \\
&\quad\times[\pi^{1/2}\operatorname{erfc}(b) + [(-1/b) + (1/2b^3) - (3/4b^5) \\
&\quad + (15/8b^7)]\exp(-b^2)] \\
&\quad + (\pi^5/168V_{\mathrm{d}})w^7T_1 - (\pi^5/240)w^{10}T_0.
\end{aligned}
$$

3.4.12. Numerical illustrations

Consider the case of the sodium chloride crystal structure (a face-centred cubic structure) as a simple example for evaluation of the Coulombic sum. The sodium ion can be taken at the origin, and the chloride ion halfway along an edge of the unit cell. The results can easily be generalized for this structure type by using the unit-cell edge length, a, as a scaling constant.

First, consider the nearest neighbours. Each sodium and each chloride ion is surrounded by six ions of opposite sign at a distance of $a/2$. The Coulombic energy for the first coordination sphere is $-(1/2)(12)(2/a)(1389.3654) = -16672.385/a$ kJ mol^{-1}. Table 3.4.2.1 shows that the converged value of the lattice energy is $-4855.979/a$. Thus the nearest-neighbour energy is over three times more negative than the total lattice energy. In the second coordination sphere each ion is surrounded by 12 similar ions at a distance of $a/2^{1/2}$. The energy contribution of the second sphere is $+(1/2)(24)(2^{1/2}/a)(1389.3654) = +23578.313/a$. Thus, major cancellation occurs and the net energy for the first two coordination spheres is $+6905.928/a$ which actually has the wrong sign for a stable crystal. The third coordination sphere again makes a negative contribution. Each ion is surrounded by eight ions of opposite sign at a distance of $a/3^{1/2}$. The energy contribution is $-(1/2)(16)(3^{1/2}/a)(1389.3654) = -19251.612/a$, now giving a total so far of $-12345.684/a$. In the fourth coordination sphere each ion is surrounded by six others of the same sign at a distance of a. The energy contribution is $+(1/2)(12)(1/a)(1389.3654) = +8336.19/a$ to yield a total of $-4009.491/a$.

It is seen immediately by examining the numbers that the Coulombic sum is converging very slowly in direct space. Madelung (1918) devised a method for accurate evaluation of the sodium chloride lattice sum. However, his method is not generally applicable for more complex lattice structures. Evjen (1932) emphasized the importance of summing over a neutral domain, and replaced the sum with an integral outside of the first few shells of nearest neighbours. But the method of Ewald remained as the only completely general and accurate method of evaluating the Coulombic sum for a general lattice. Although it was derived in a somewhat different way, Ewald's method is equivalent to accelerated convergence for the special case of $n = 1$.

In the reciprocal lattice of sodium chloride only points with indices (hkl) all even or all odd are permitted by the face-centred symmetry. The reciprocal cell has edge length $1/a$ and the reciprocal-axis directions coincide with the direct-lattice axis directions. The closest points to the origin are the eight (111) forms at a distance of $(1/a)/3^{1/2}$. For sodium chloride this distance is 0.3078 Å^{-1}.

Table 3.4.12.1 shows the effect of convergence acceleration on the direct-space portion of the $(n = 1)$ sum for the sodium chloride structure. The constant term $-w\sum q_j^2$ is included in the values given. This constant term is always large if w is not zero; for instance, when $w = 0.1$ this term is −277.872 (Table 3.4.12.2). For $w = 0.1$ the reciprocal-lattice sum is zero to six figures. Thus, only the direct sum (plus the constant term) is needed, evaluated out to 20 Å in direct space, to obtain six-figure accuracy. As shown in Table 3.4.2.1 above, the same summation effort without the use of accelerated convergence gave 8% error, or only

Table 3.4.12.3. *Accelerated-convergence results for the dispersion sum (n = 6) of crystalline benzene (kJ mol^{-1}, Å); the figures shown are the direct-lattice sum plus the two constant terms*

Limit	$w = 0.1$	$w = 0.15$	$w = 0.2$	$w = 0.3$	$w = 0.4$
6.0	−73.452	−77.761	−79.651	−61.866	76.645
8.0	−79.029	−80.374	−80.256	−61.870	76.645
10.0	−80.217	−80.571	−80.265	−61.870	
12.0	−80.527	−80.585	−80.265		
14.0	−80.578	−80.585			
16.0	−80.588				
18.0	−80.589				
20.0	−80.589				

Table 3.4.12.4. *The reciprocal-lattice results (kJ mol^{-1}, Å) for the dispersion sum (n = 6) of crystalline benzene*

Limit	$w = 0.1$	$w = 0.15$	$w = 0.2$	$w = 0.3$	$w = 0.4$
0.0	−5.547	−16.706	−32.326	−43.681	80.947
0.3	0.000	−0.004	−0.321	−16.792	−117.106
0.4	0.000	−0.004	−0.324	−18.656	−152.651
0.5			−0.324	−18.716	−155.940
0.6				−18.719	−157.102
0.7				−18.719	−157.227
0.8					−157.233
0.9					−157.234
1.0					−157.234

Table 3.4.12.5. *Approximate time (s) required to evaluate the dispersion sum (n = 6) for crystalline benzene within 0.001 kJ mol^{-1} truncation error*

w	Direct terms	Reciprocal terms	Time, direct	Time, reciprocal	Total time
0.0	(not yet converged at 20 Å summation limit)				>107
0.1	15904	0	77	0	77
0.15	4718	34	23	6	29
0.2	2631	78	13	14	27
0.3	1313	246	7	46	53
0.4	524	804	3	149	152

slightly better than one-figure accuracy. The accelerated-convergence technique therefore yielded nearly five orders of magnitude improvement in accuracy, even without evaluation of the reciprocal-lattice sum.

The column showing $w = 0.15$ shows an example of how the reciprocal-lattice sum can also be neglected if lower accuracy is required. Table 3.4.12.2 shows that the reciprocal-lattice sum is still only 0.003. But now the direct-lattice sum only needs to be evaluated out to 14 Å, with further savings in calculation effort. For w values larger than 0.15 the reciprocal sum is needed. For $w = 0.4$ this sum must be evaluated out to 0.8 Å^{-1} to obtain six-figure accuracy.

Table 3.4.12.3 illustrates an application for the $(n = 6)$ dispersion sum. When $w = 0.1$ Å^{-1} five figures of accuracy can be obtained without consideration of the reciprocal sum. The direct sum is required out to 18 Å. If $w = 0.15$, better than four-figure accuracy can still be obtained without evaluating the reciprocal-lattice sum. In this case, the direct lattice needs to be summed only to 12 Å, and there is a saving of an order of magnitude in the length of the calculation. As with the Coulombic sum, if w is greater than 0.15 the reciprocal-lattice summation is needed; Table 3.4.12.4 shows the values.

The time required to obtain a lattice sum of given accuracy will vary depending on the particular structure considered and of course on the computer and program which are used. An example of timing for the benzene dispersion sum is given in Table 3.4.12.5 for the *PCK83* program (Williams, 1984) running on a VAX-11/750 computer. In this particular case direct terms were evaluated at a rate of about 200 terms s^{-1} and reciprocal terms, being a sum themselves, were evaluated at a slower rate of about 5 terms s^{-1}.

Table 3.4.12.5 shows the time required for evaluation of the dispersion sum using various values of the convergence constant, w. The timing figures show that there is an optimum choice for w; for the *PCK83* program the optimum value indicated is 0.15–0.2 Å^{-1}. In the program of Pietila & Rasmussen (1984) values in the range 0.15–0.2 Å^{-1} are also suggested. For the *WMIN* program (Busing, 1981) a slightly higher value of 0.25 Å^{-1} is suggested. Trial calculations can be used to determine the optimum value of w for the situation of a particular crystal structure, program and computer.

References

Arfken, G. (1970). *Mathematical Methods for Physicists,* 2nd ed. New York: Academic Press.

Bertaut, E. F. (1952). *L'énergie électrostatique de réseaux ioniques. J. Phys. (Paris)*, **13**, 499–505.

Bertaut, E. F. (1978). *The equivalent charge concept and its application to the electrostatic energy of charges and multipoles. J. Phys. (Paris)*, **39**, 1331–1348.

Busing, W. R. (1981). *WMIN, a computer program to model molecules and crystals in terms of potential energy functions.* Oak Ridge National Laboratory Report ORNL-5747. Oak Ridge, Tennessee 37830, USA.

Cummins, P. G., Dunmur, D. A., Munn, R. W. & Newham, R. J. (1976). *Applications of the Ewald method. I. Calculation of multipole lattice sums. Acta Cryst.* A**32**, 847–853.

Davis, P. J. (1972). *Gamma function and related functions. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz & I. A. Stegun, pp. 260–262. London, New York: John Wiley. [Reprint, with corrections of 1964 Natl Bur. Stand. publication.]

DeWette, F. W. & Schacher, G. E. (1964). *Internal field in general dipole lattices. Phys. Rev.* **137**, A78–A91.

Evjen, H. M. (1932). *The stability of certain heteropolar crystals. Phys. Rev.* **39**, 675–694.

Ewald, P. P. (1921). *Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys. (Leipzig)*, **64**, 253–287.

Fortuin, C. M. (1977). *Note on the calculation of electrostatic lattice potentials. Physica (Utrecht)*, **86A**, 574–586.

Glasser, M. L. & Zucker, I. J. (1980). *Lattice sums. Theor. Chem. Adv. Perspect.* **5**, 67–139.

Hastings, C. Jr (1955). *Approximations for digital computers.* New Jersey: Princeton University Press.

Madelung, E. (1918). *Das elektrische Feld in Systemen von regelmässig angeordneten Punktladungen. Phys. Z.* **19**, 524–532.

Massidda, V. (1978). *Electrostatic energy in ionic crystals by the planewise summation method. Physica (Utrecht)*, **95B**, 317–334.

Nijboer, B. R. A. & DeWette, F. W. (1957). *On the calculation of lattice sums. Physica (Utrecht)*, **23**, 309–321.

Pietila, L.-O. & Rasmussen, K. (1984). *A program for calculation of crystal conformations of flexible molecules using convergence acceleration. J. Comput. Chem.* **5**, 252–260.

Widder, D. V. (1961). *Advanced Calculus*, 2nd ed. New York: Prentice-Hall.

Williams, D. E. (1971). *Accelerated convergence of crystal lattice potential sums. Acta Cryst.* A**27**, 452–455.

Williams, D. E. (1984). *PCK83, a crystal molecular packing analysis program.* Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA.

Williams, D. E. (1989*a*). *Accelerated convergence treatment of R^{-n} lattice sums. Crystallogr. Rev.* **2**, 3–23.

Williams, D. E. (1989*b*). *Accelerated convergence treatment of R^{-n} lattice sums. Corrections. Crystallogr. Rev.* **2**, 163–166.

International Tables for Crystallography (2010). Vol. B, Chapter 3.5, pp. 458–481.

3.5. Extensions of the Ewald method for Coulomb interactions in crystals

By T. A. Darden

3.5.1. Introduction

High-precision single-crystal X-ray structural analysis of small organic molecules, yielding the space group, the unit-cell parameters and the fractional coordinates of the atoms making up the molecule(s) in the asymmetric unit, has become a routine matter as long as crystals of sufficient quality can be obtained. The thermodynamic stability of the crystal, as described by the enthalpy of sublimation ΔH_{sub}, can also be determined experimentally (although not always to high precision). Theoretical models for calculating intermolecular interaction energies can be used to connect the crystal structure to the molar enthalpy of sublimation using the relationship

$$\Delta H_{sub}(0\ \mathrm{K}) = -E_{lattice},$$

where the lattice or packing energy $E_{lattice}$ is the total (molar) intermolecular interaction energy between all the molecules in the crystal, which are treated as rigid entities with zero-point energies of intra- and intermolecular vibrations neglected. Connection to experimentally accessible heats of sublimation at higher temperatures involves thermodynamic corrections. Methods for calculating thermodynamic quantities of solids are discussed in Gavezzotti (2002*a*) and (in more detail) in Frenkel & Smit (2002).

Thus, given a parameterized intermolecular potential-energy function, or if computationally affordable a first-principles approach such as density-functional theory (or preferably, when it becomes feasible for crystals, a good-quality post-Hartree–Fock potential-energy surface that describes dispersion interactions), one can sum the intermolecular energies to obtain the lattice energy as a function of the above parameters defining the crystal structure. If such an energy function is used together with a method for systematic search of the crystal-structure parameters, one could in principle predict the minimum-lattice-energy crystal structure for a rigid organic molecule. To extend this approach to flexible molecules one would need to minimize the sum of the intramolecular energy plus the lattice energy. If the experimental crystal structure corresponds to the thermodynamic minimum-energy structure (*i.e.* it is not a metastable state determined by crystal-growth kinetics), one could in principle predict the experimental crystal structure of an organic compound through this minimization protocol. Moreover, one could ideally predict the additional metastable forms of the crystal.

Prediction of the structure of crystals of an organic molecule from its molecular structure is a difficult problem that has been compared to the protein folding problem (Dunitz, 2003; Dunitz & Scheraga, 2004). Like the protein folding problem, a solution of the crystal prediction problem has significant practical ramifications. A compound is often polymorphic, that is it has more than one crystal structure, and it may be difficult to characterize the conditions under which a particular crystal structure is formed (Dunitz & Bernstein, 1995). Polymorphs may have very different physical properties. An obvious example is diamond *versus* graphite, but other commercially important examples include food additives, various solid forms of explosives and the bioavailability of various forms of a drug such as ritonavir (Chemburkar *et al.*, 2000). A method for predicting the possible crystal structures of the compound, and ideally for predicting the dominant crystal structure given the experimental conditions, would thus be very valuable. Note that due to the subtleties of crystallization, the lowest-free-energy polymorph at given temperature and pressure may not be the likeliest to form. The kinetics of growth of microcrystals may largely determine which low-energy polymorph appears (Dunitz, 2003). However, it is generally agreed that accurate calculation of the relative free energy of polymorphs is a prerequisite for predicting crystal structures.

To assess progress towards solving this latter problem, a series of blind tests of crystal-structure prediction has been undertaken (Day, Motherwell, Ammon *et al.*, 2005). The results of these tests have highlighted the need for continued improvements in sampling methods and intermolecular energy potentials. Since extensive sampling of the crystal-structure parameters is necessary [between 10^4 and 10^5 starting structures, each followed by parameter minimizations (Price & Price, 2005)], there is a trade-off in the computational cost *versus* accuracy of the intermolecular energy functions used. Calculating the work of transforming between polymorphs is yet more ambitious in terms of sampling. Consequently empirical force fields are likely to be needed for the near term at least.

In the remainder of the introduction we outline some of the approaches to empirical potentials used in the calculation of the lattice energy, and then, motivated by these developments, discuss techniques for efficient summation of the electrostatic and other slow-decaying interaction terms that occur in these potential functions.

Methodological developments in the intermolecular force fields used in crystal-structure prediction from early times to the present state of the art have been reviewed (Price & Price, 2005). Until recently, these force fields were made up of atom–atom interactions. The earliest involved only repulsion and dispersion, usually in the 'exp-6' form

$$U^{MN} = U^{MN}_{rep} + U^{MN}_{disp} = \sum_{i\in M, j\in N} A_{ij}\exp(-B_{ij}r_{ij}) - C_{i,j}r_{ij}^{-6},$$

where U^{MN} denotes the intermolecular potential energy between molecules M and N and r_{ij} is the distance between atoms $i \in M$ and $j \in N$. Sometimes the exponential form in the above equation is replaced by a simpler power law, as in the Lennard–Jones potential. As was pointed out by Dunitz (2003), in comparison with more sophisticated force fields, this repulsion–dispersion form readily allows analysis of the significance of particular atom–atom interactions, since the interactions are short-ranged and thus can be localized. That is, the r^{-6} form of the attractive dispersion energy means that interaction energies are halved for every 12% increase in distance. In contrast, introduction of long-range Coulombic interactions not only entails subtleties in lattice summation (the subject of this contribution), but greatly complicates the assignment of 'key' atom–atom interactions. Gavezzotti and Fillipini systematically explored the use of the exp-6 potential in fitting organic crystal structures with and without hydrogen-bond interactions (Gavezzotti & Fillipini, 1994). They were surprisingly successful in accounting for weak hydrogen bonding in this way, but selective use of point charges improved the directionality of the potential. Earlier, Williams derived exp-6 parameters for the atoms C, H, N, O, Cl, F and polar H for use in organic crystal structures, but found it necessary (Williams & Cox, 1984) to supplement these with selected point charges, both atomic and at off-atom sites. Price and

Copyright © 2010 International Union of Crystallography

co-workers have used a more sophisticated electrostatic model based on distributed multipole analysis (DMA) with multipoles up to hexadecapole order positioned at atom centres (Stone, 1996). A similar approach has been adopted by Mooji *et al.*, with the addition of dipole polarizabilities to account for intermolecular induction energy (Mooji *et al.*, 1999). The AMOEBA force field has similar capabilities (Ren & Ponder, 2003).

Recently, a systematic study (Day *et al.*, 2004; Day, Motherwell & Jones, 2005) of the effect of the electrostatic model on crystal-structure prediction for rigid organic molecules was carried out. The exp-6 parameters from the W99 force field (Williams, 2001) were used in conjunction with point charges derived either from a bond-increment model or electrostatic potential (ESP) fit, or with a DMA model as used by Price and co-workers. Importantly, they found a systematic and significant improvement in prediction success going from the simple bond-increment model to the ESP charges to point multipoles. Similar results were found by Broderson *et al.* (2003). This is a positive indicator for the use of theory in attacking this difficult prediction problem. On the other hand, Broderson *et al.* found that more sophisticated electrostatic models did not help in predicting the packing of flexible molecules. As discussed by Price & Price (2005), crystal-packing energies of flexible molecules are more problematic for at least two reasons. First the intramolecular energy must be fully compatible with the intermolecular energy in order that energy deformation from the gas-phase minimum be on the same scale as the concomitant gain in packing energy. Secondly, the conformational dependence of the electrostatic model must be accurately accounted for, or any gains from a more sophisticated model are immediately negated. Existing force fields are unable to meet these challenges, which may explain the lower success rate for predicting crystal structures of flexible molecules. Price has developed a methodology for flexible molecules wherein intramolecular conformations are systematically explored. For each conformer studied, the gas-phase intramolecular energy and DMA electrostatic model are calculated using quantum-chemical methods. The optimal packing of that conformer is then obtained by treating it as a rigid molecule, and the resulting minimum lattice energy is added to its intramolecular energy to give its total energy. The total energy for that conformer is then compared with that of the other conformers to predict the optimal conformation and packing of the flexible molecule. This approach has proved successful in a number of cases, for example with the successful prediction of a novel second polymorph of 1-hydroxy-7-azabenzotriazole (Nowell *et al.*, 2006).

The above electrostatic models are all derived from gas-phase quantum-chemical calculations. A different line of development has been the parameterization of electrostatic models from condensed-phase experimental or theoretical electron densities, which should better represent the polarization due to the crystal environment. This is accomplished using aspherical multipole refinement of experimental X-ray charge densities (Destro *et al.*, 2000) or of theoretical charge densities from quantum-chemical calculations carried out under the space-group symmetry (Dovesi *et al.*, 2005). Crystal structure analysis is usually carried out assuming the independent-atom model (IAM), where the individual atomic densities are modelled as those of the spherically symmetric isolated atoms, with deviations in the experimental density being fitted to anisotropic temperature factors modelling the thermal motion of the nuclei. However, the IAM breaks down for data of sufficiently high quality, and the residual or deformation density is then fitted by a linear combination of atom-centred density basis elements resembling the well known Slater atomic basis sets from quantum chemistry (*i.e.* solid harmonics times a decaying exponential in the distance r from the atom). Technical details are provided in Coppens (1997) and Chapter 1.2 of this volume. The theoretical density can be similarly approached as a sum of pro-molecule density (superposition of isolated atomic densities in the molecular geometry) plus deformation density modelled by a sum of atomic Slater-like density elements. Until recently, the theoretical densities were thought to be of lower quality than the experimental densities, but continued improvements in the level of theory that can be applied have resulted in much closer agreement (Coppens & Volkov, 2004). Coppens and co-workers as well as others (Spackman *et al.*, 1988) have applied these modelled densities in studies of intermolecular electrostatic energies. The Slater-like functions are typically difficult to work with in this context, so multipole approximations are used except at close range, where numerical integration is used instead (Volkov *et al.*, 2004).

A similar density-based approach using gas-phase theoretical density was developed previously (Gavezzotti, 2002*b*). Gavezzotti splits the density into numerical 'pixels' (small cubes of volume) within which the density is considered constant, allowing in principle exact numerical calculation of intermolecular electrostatic interactions. He sums the pixel–pixel interactions discretely, using a hierarchical approach to increase efficiency. The electrostatic interactions are supplemented by induction and repulsion terms, also modelled by pixel-based interactions. These developments, in particular the large size of the positive and negative components of the electrostatic interactions treated in terms of actual electron density, have caused a reassessment of the idea that intermolecular interactions can be accurately decomposed even in principle into localized atom–atom interactions (Dunitz & Gavezzotti, 2005).

Cisneros *et al.* have also begun implementing an intermolecular potential based on the explicit modelling of gas-phase theoretical electron density, using the methodology of auxiliary basis set fitting (Cisneros *et al.*, 2005). The density is expanded into atom and off-atom sites using Gaussian density functions similar to those used as atomic basis sets in quantum chemical codes. More recently, other components of the potential such as induction, charge transfer and non-isotropic site–site repulsion based on density overlap have been implemented (Piquemal *et al.*, 2006) and efficient calculation of terms in periodic boundary conditions has also been implemented (Cisneros *et al.*, 2006), generalizing previous developments for point multipoles (Sagui *et al.*, 2004).

Summarizing the above developments, it appears that accurate evaluation of crystal lattice energies necessitates the use of complex electrostatic interactions beyond the spherical-atom approximation, and ultimately including an account of penetration effects through explicit modelling of electron density. The electrostatic lattice energies are differences between large positive and negative contributions that are long-ranged, necessitating an accurate treatment of their lattice summation. Dispersion interactions, modelled by higher-order inverse powers of distance, are short-ranged but their calculation can be greatly accelerated through lattice-summation techniques.

Ewald summation in ideal infinite crystals, for interactions that depend on inverse powers of site–site distance, is covered in Chapter 3.4 of this volume, where there is also a numerical demonstration of the advantages of the Ewald approach over direct numerical summation. In this contribution we focus on the lattice-summation problem for large but finite crystals, highlighting the role of summation order and leading to a derivation of the surface term and polarization response. We also consider extensions to Coulombic interactions between Gaussian density elements, both spherical and Hermite (multipolar), and methods to further accelerate the lattice summation to (almost) linear scaling.

These developments are organized into three sections. The first section treats lattice summation of point charges. We begin with definitions and an explanation of the origin of the conditional convergence problem for these sums, anticipating the nature of the results to follow. Following that we provide for the benefit of less mathematically oriented readers the standard physics-based derivation of the Ewald sum for point charges, minus any

accounting for the order of convergence or resulting surface term. This is followed by a more careful derivation focused on the order of summation, including results for higher-order inverse powers as well as the Coulomb interactions, finishing with a discussion of the surface term, the polarization response to an external dielectric and the pressure calculation. Although pitched at a higher mathematical level, we have tried to provide many details in the derivations unavailable elsewhere for the benefit of the reader. The alert reader will note that the key identity, equation (3.5.2.19), that separates $1/r^n$ into short-ranged and long-ranged components, is the same as that used Chapter 3.4 of this volume, highlighting the essential similarity of our approach. In the second section we generalize these point-charge results to lattice sums of interacting spherical Gaussian and Hermite Gaussian-based densities. The third and final section provides a discussion of two methods that greatly increase the numerical efficiency of these lattice sums in the case of large unit cells by utilizing the fast Fourier transform. Again we have tried to provide mathematical details not available elsewhere.

3.5.2. Lattice sums of point charges

We begin by discussing an idealized infinite crystal C, made up of point charges.

3.5.2.1. Basic quantities

The periodicity of an idealized crystal C is specified by lattice basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, which are linearly independent as vectors and thus form a basis for the usual vector space of points in three dimensions. The conjugate reciprocal vectors $\mathbf{a}_\alpha^*$ are defined by the relations $\mathbf{a}_\alpha^* \cdot \mathbf{a}_\beta = \delta_{\alpha\beta}$ (the Kronecker delta) for $\alpha, \beta = 1, 2, 3$. Thus for example $\mathbf{a}_1^*$ is given by $\mathbf{a}_1^* = (\mathbf{a}_2 \times \mathbf{a}_3)/[\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)]$. General lattice vectors $\mathbf{n}$ are given by integral combinations of the lattice basis vectors:

$$\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{3.5.2.1}$$

where n_1, n_2, n_3 are integers. General reciprocal-lattice vectors $\mathbf{m}$ are given by integral combinations of the reciprocal-lattice basis vectors:

$$\mathbf{m} = m_1\mathbf{a}_1^* + m_2\mathbf{a}_2^* + m_3\mathbf{a}_3^*. \tag{3.5.2.2}$$

An arbitrary point $\mathbf{r}$ in the crystal ($\mathbf{r}$ also denotes the vector from the origin of coordinates to the point) is given by

$$\mathbf{r} = s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + s_3\mathbf{a}_3,$$

where s_1, s_2, s_3 are the fractional coordinates of the point. From the above equation and the defining relations for the reciprocal basis we see that $s_\alpha = \mathbf{r} \cdot \mathbf{a}_\alpha^*$ for $\alpha = 1, 2, 3$. The unit cell U is the set of points $\mathbf{r}$ with associated fractional coordinates s_1, s_2, s_3 satisfying $-\frac{1}{2} \leq s_\alpha \leq \frac{1}{2}, \alpha = 1, 2, 3$. The idealized infinite crystal C is generated by the union of all periodic translations of the unit cell U, using the set of general lattice vectors $\mathbf{n}$. A point charge q at position $\mathbf{r}$ in U has periodic 'image' charges $q_\mathbf{n}$ at positions $\mathbf{r}_\mathbf{n} = \mathbf{r} + \mathbf{n}$ for all lattice vector $\mathbf{n}$. The periodic image of U, denoted $U_\mathbf{n}$, is given by the periodic images $q_\mathbf{n}$ of all point charges q in U. Thus, the crystal is made up of the union of images $U_\mathbf{n}$, including $U = U_\mathbf{0}$, or

$$C = \bigcup_\mathbf{n} U_\mathbf{n}.$$

A point charge q_i in U at position $\mathbf{r}_i$ interacts with other unit-cell charges $q_j, j \neq i$ at positions $\mathbf{r}_j$ as well as with all of their periodic images $q_{j\mathbf{n}}$ at positions $\mathbf{r}_{j\mathbf{n}} = \mathbf{r}_j + \mathbf{n}$ for all lattice vectors $\mathbf{n}$. It also interacts with its own periodic images $q_{i\mathbf{n}}$ at $\mathbf{r}_{i\mathbf{n}}$ for all nonzero lattice vectors $\mathbf{n}$, that is all $\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ with $(n_1, n_2, n_3) \neq (0, 0, 0)$. Suppose the unit cell U is made up of N point charges $q_1, \ldots, q_N$ at positions $\mathbf{r}_1, \ldots, \mathbf{r}_N$. Then the electrostatic energy of U, given by the sum of the Coulombic interactions of charges in U with each other and with the rest of C, is written as

$$\begin{aligned} E(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= E(\mathbf{r}_1, \ldots, \mathbf{r}_N) \\ &= \tfrac{1}{2}{\sum_\mathbf{n}}' \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{n}|}, \end{aligned} \tag{3.5.2.3}$$

where here and below $\underline{\mathbf{r}}^{\{\mathbf{N}\}}$ denotes the set of points $\{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ and where the outer sum on the right is over the general lattice vectors $\mathbf{n}$, the prime indicating that terms with $i = j$ and $\mathbf{n} = \mathbf{0}$ are omitted. The treatment of this infinite sum requires some care. The energy diverges unless the unit cell is neutral (*i.e.* unless $\sum_{i=1}^N q_i = 0$). When U is neutral, the sum, if carried out in a naive fashion, can still converge quite slowly (see Chapter 3.4). In addition, the convergence is *conditional*, that is the resulting energy can vary depending on the order in which the sum is carried out! To better understand these phenomena, it is helpful to re-express the above energy. Let $E(U, U)$ denote the Coulomb interaction energy of U with itself and $E(U, U_\mathbf{n}), \mathbf{n} \neq \mathbf{0}$ the Coulomb interaction energy of U with $U_\mathbf{n}$, that is

$$E(U, U) = \sum_{i<j} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$$

and

$$E(U, U_\mathbf{n}) = \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{n}|} = \sum_{q_i \in U} \sum_{q_{j\mathbf{n}} \in U_\mathbf{n}} \frac{q_i q_{j\mathbf{n}}}{|\mathbf{r}_i - \mathbf{r}_{j\mathbf{n}}|}.$$

Then we can abbreviate equation (3.5.2.3) by

$$E(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) = E(U, U) + \tfrac{1}{2}\sum_{\mathbf{n}\neq\mathbf{0}} E(U, U_\mathbf{n}). \tag{3.5.2.4}$$

Note that for large $\mathbf{n}$ (that is, $\mathbf{n}$ with large norm $|\mathbf{n}|$) $E(U, U_\mathbf{n})$ can be approximated by a Taylor expansion. Let $\bar{\mathbf{r}}$ denote the centroid of points in U, given by $\bar{\mathbf{r}} = (1/N)\sum_{i=1}^N \mathbf{r}_i$, and let $\bar{\mathbf{r}}_\mathbf{n}$ denote the centroid of $U_\mathbf{n}$. Then, expanding in terms of $\mathbf{r}_i - \bar{\mathbf{r}} = \Delta\mathbf{r}_i$ and $\mathbf{r}_{j\mathbf{n}} - \bar{\mathbf{r}}_\mathbf{n} = \mathbf{r}_j - \bar{\mathbf{r}} = \Delta\mathbf{r}_j$ (which are small compared to $\bar{\mathbf{r}} - \bar{\mathbf{r}}_\mathbf{n} = -\mathbf{n}$) we have

$$\begin{aligned}\frac{1}{|\mathbf{r}_i-\mathbf{r}_{j\mathbf{n}}|} &= \frac{1}{|(\bar{\mathbf{r}}-\bar{\mathbf{r}}_{\mathbf{n}})+(\mathbf{r}_i-\bar{\mathbf{r}})-(\mathbf{r}_{j\mathbf{n}}-\bar{\mathbf{r}}_{\mathbf{n}})|}\\ &= \frac{1}{|\mathbf{n}-(\Delta\mathbf{r}_i-\Delta\mathbf{r}_j)|}\\ &\simeq \frac{1}{|\mathbf{n}|}+\sum_{\alpha=1}^{3}\frac{\mathbf{n}_\alpha}{|\mathbf{n}|^3}(\Delta\mathbf{r}_i-\Delta\mathbf{r}_j)_\alpha\\ &\quad+\tfrac{1}{2}\sum_{\alpha,\beta=1}^{3}\frac{(3\mathbf{n}_\alpha\mathbf{n}_\beta-\delta_{\alpha\beta}\mathbf{n}^2)}{|\mathbf{n}|^5}\\ &\quad\times(\Delta\mathbf{r}_i-\Delta\mathbf{r}_j)_\alpha(\Delta\mathbf{r}_i-\Delta\mathbf{r}_j)_\beta\\ &\quad+\ldots.\end{aligned}$$

Let $Q=\sum_{i=1}^{N}q_i$ and $\mathbf{D}=\sum_{i=1}^{N}q_i\Delta\mathbf{r}_i$ denote the net charge and dipole moment of the unit cell U, and let $\mathbf{G}_{\alpha\beta}=\sum_{i=1}^{N}q_i\Delta\mathbf{r}_{i\alpha}\Delta\mathbf{r}_{i\beta}$ denote the quadrupole moment of the unit cell. Using the expansion in the above equation, we can Taylor expand the interaction energy $E(U,U_{\mathbf{n}})$ as

$$\begin{aligned}E(U,U_{\mathbf{n}}) &= \sum_{i=1}^{N}\sum_{j=1}^{N}q_iq_j\frac{1}{|\mathbf{n}-(\Delta\mathbf{r}_i-\Delta\mathbf{r}_j)|}\\ &\simeq \frac{1}{|\mathbf{n}|}Q^2\\ &\quad+\sum_{\alpha,\beta=1}^{3}\frac{(3\mathbf{n}_\alpha\mathbf{n}_\beta-\delta_{\alpha\beta}\mathbf{n}^2)}{|\mathbf{n}|^5}(Q\mathbf{G}_{\alpha\beta}-\mathbf{D}_\alpha\mathbf{D}_\beta)\\ &\quad+\ldots.\end{aligned}\tag{3.5.2.5}$$

Note that the first-order term in the expansion has dropped out, since the dipole moments $\mathbf{D}$ of U and $U_{\mathbf{n}}$ are the same. If $Q\neq 0$ then for large $|\mathbf{n}|$ the leading term in the Taylor expansion of $E(U,U_{\mathbf{n}})$ is $Q^2/|\mathbf{n}|$. Since $\int\int\int(1/|\mathbf{r}|)\,\mathrm{d}^3\mathbf{r}=4\pi\int_0^\infty r\,\mathrm{d}r=\infty$, then using a standard integral comparison $\sum_{\mathbf{n}\neq\mathbf{0}}1/|\mathbf{n}|$ is infinite, and so the resulting electrostatic energy of U diverges. If $Q=0$ we see that the leading term in the expansion of $E(U,U_{\mathbf{n}})$ is given by $-\sum_{\alpha,\beta}[(3\mathbf{n}_\alpha\mathbf{n}_\beta-\delta_{\alpha\beta}\mathbf{n}^2)/|\mathbf{n}|^5]\mathbf{D}_\alpha\mathbf{D}_\beta$, that is a quadratic expression in the dipole moment $\mathbf{D}$ times a prefactor that decays like $1/|\mathbf{n}|^3$. Since $\int\int\int 1/|\mathbf{r}|^3\,\mathrm{d}^3\mathbf{r}=4\pi\int_0^\infty 1/r\,\mathrm{d}r=\infty$, $\sum_{\mathbf{n}\neq\mathbf{0}}1/|\mathbf{n}|^3$ also diverges (logarithmically), and so the electrostatic energy of U cannot converge absolutely in this case either. Hence, if it converges at all, it does so *conditionally*, that is the energy of U depends on the order in which we sum the contributions from image cells $U_{\mathbf{n}}$. If $Q=0$ and $\mathbf{D}=\mathbf{0}$, the leading term in the Taylor expansion of $E(U,U_{\mathbf{n}})$ depends on the quadrupole moment of the unit cell and decays like $1/|\mathbf{n}|^5$ for large $|\mathbf{n}|$, and thus since $\sum_{\mathbf{n}\neq\mathbf{0}}1/|\mathbf{n}|^5$ does converge, the electrostatic energy of U converges absolutely. From these considerations we might hope to express the energy of a neutral unit cell U in terms of an absolutely convergent sum plus a term quadratic in the unit-cell dipole moment $\mathbf{D}$ which also depends on the method of summation. As we will see below this is precisely what the Ewald summation method gives us.

3.5.2.2. Ewald sum: first derivation

In this section we give the standard Ewald-sum derivation found in textbooks such as Kittel (1986). To do this we need two preliminary results, namely the electrostatic potential due to an isolated Gaussian charge distribution having total charge of one, and then that due to a periodic array of such distributions.

First we derive the electrostatic potential $\Psi(\mathbf{r})$ due to an isolated Gaussian density $\rho(\mathbf{r})$ given by

$$\rho(\mathbf{r})=(\alpha/\pi)^{3/2}\exp(-\alpha\mathbf{r}^2).\tag{3.5.2.6}$$

Owing to the spherical symmetry of $\rho(\mathbf{r})$ we can define the one-dimensional function $\psi(r)$ by $\psi(r)=\Psi(\mathbf{r})$ for any $\mathbf{r}$ with $|\mathbf{r}|=r$. We next invoke Gauss' law (Eyges, 1980) under spherical symmetry to evaluate the electric field $\mathbf{E}(\mathbf{r})$ due to ρ at any distance $r=|\mathbf{r}|$. In the case of spherical symmetry Gauss' law states that the electric field at a point $\mathbf{r}$, due to a spherically symmetric charge distribution centred at the origin, is the same as if all the charge of the distribution from the origin out to $|\mathbf{r}|$ were collected into a point charge at the origin. More specifically, if $\mathbf{S}$ is the sphere of radius $r=|\mathbf{r}|$ with surface $\partial\mathbf{S}$, Gauss' law states

$$(1/4\pi)\int_{\partial\mathbf{S}}\mathbf{E}\,\mathrm{d}\mathbf{s}=\int_{\mathbf{S}}\rho(\mathbf{u})\,\mathrm{d}^3\mathbf{u},$$

and thus, defining the scalar quantity $E(r)=|\mathbf{E}(\mathbf{r})|$, we get

$$\begin{aligned}E(r) &= \frac{1}{r^2}\int_{\mathbf{S}}\rho(\mathbf{u})\,\mathrm{d}^3\mathbf{u}\\ &= \frac{1}{r^2}\left(\frac{\alpha}{\pi}\right)^{3/2}\int_{u=0}^{r}4\pi u^2\exp(-\alpha u^2)\,\mathrm{d}u\\ &= \frac{\mathrm{erf}(\alpha^{1/2}r)}{r^2}-\frac{2(\alpha/\pi)^{1/2}\exp(-\alpha r^2)}{r}\end{aligned}$$

and thus, since (again by symmetry) $E(r)=-(\mathrm{d}/\mathrm{d}r)\psi(r)$, we can verify that

$$\psi(r)=\frac{\mathrm{erf}(\alpha^{1/2}r)}{r}+C,\tag{3.5.2.7}$$

where C is an arbitrary constant of integration. We obtain a potential with the desired asymptotic behaviour, *i.e.* $\psi(r)\to 0$ as $r\to\infty$, by setting $C=0$. Note then that $\psi(r)\simeq 1/r$ for large r as expected.

Next we derive the electrostatic potential $\Phi^{\mathrm{P}}(\mathbf{r})$ due to a periodic array of Gaussian charge distributions, each of the form just studied. That is, for points $\mathbf{r}$ in the unit cell U, define the periodic density $\rho^{\mathrm{P}}(\mathbf{r})$ by

$$\rho^{\mathrm{P}}(\mathbf{r})=\sum_{\mathbf{n}}\rho(\mathbf{r}+\mathbf{n}),$$

where the density ρ is defined in equation (3.5.2.6) and the sum is over all general lattice vectors $\mathbf{n}$, defined in equation (3.5.2.1). Since ρ^{P} is periodic, *i.e.* $\rho^{\mathrm{P}}(\mathbf{r}+\mathbf{n}_0)=\rho(\mathbf{r})$ for any general lattice vector $\mathbf{n}_0$, we can expand it in a Fourier series over the unit cell:

$$\rho^{\mathrm{P}}(\mathbf{r})=\sum_{\mathbf{m}}C_{\mathbf{m}}\exp(2\pi i\mathbf{m}\cdot\mathbf{r}),\tag{3.5.2.8}$$

where the sum is over all reciprocal-lattice vectors $\mathbf{m}$, defined in equation (3.5.2.2), and the Fourier coefficient $C_{\mathbf{m}}$ is given by

$$
\begin{aligned}
C_{\mathbf{m}} &= \frac{1}{V}\int_U \rho^{\mathrm{P}}(\mathbf{r})\exp(-2\pi i\mathbf{m}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{r} \\
&= \frac{1}{V}\sum_{\mathbf{n}}\int_U \rho(\mathbf{r}+\mathbf{n})\exp[-2\pi i\mathbf{m}\cdot(\mathbf{r}+\mathbf{n})]\,\mathrm{d}^3\mathbf{r} \\
&= \frac{1}{V}\int_{\Re^3} \rho(\mathbf{r})\exp(-2\pi i\mathbf{m}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{r} \\
&= \frac{1}{V}\left(\frac{\alpha}{\pi}\right)^{3/2}\int_{\Re^3} \exp[-\alpha(\mathbf{r}+\pi i\mathbf{m})]^2\exp(-\pi^2\mathbf{m}^2/\alpha)\,\mathrm{d}^3\mathbf{r} \\
&= \frac{1}{V}\exp(-\pi^2\mathbf{m}^2/\alpha), \qquad (3.5.2.9)
\end{aligned}
$$

where V is the volume of the unit cell U. The periodicity of ρ^{P} implies that of Φ^{P}, so we can also expand the latter in a Fourier series:

$$\Phi^{\mathrm{P}}(\mathbf{r}) = \sum_{\mathbf{m}} B_{\mathbf{m}}\exp(2\pi i\mathbf{m}\cdot\mathbf{r}). \qquad (3.5.2.10)$$

Next we recall Poisson's equation (Eyges, 1980) for Φ^{P} in terms of ρ^{P}:

$$\nabla\cdot\nabla\Phi^{\mathrm{P}}(\mathbf{r}) = -4\pi\rho^{\mathrm{P}}(\mathbf{r}). \qquad (3.5.2.11)$$

Using equations (3.5.2.10) and (3.5.2.8) we can rewrite equation (3.5.2.11) as

$$\sum_{\mathbf{m}}(-4\pi^2\mathbf{m}^2)B_{\mathbf{m}}\exp(2\pi i\mathbf{m}\cdot\mathbf{r}) = -4\pi\sum_{\mathbf{m}} C_{\mathbf{m}}\exp(2\pi i\mathbf{m}\cdot\mathbf{r}),$$

thus using the uniqueness of Fourier-series coefficients $\pi\mathbf{m}^2B_{\mathbf{m}} = C_{\mathbf{m}}$. If $\mathbf{m}\neq\mathbf{0}$ this latter equation can be solved for $B_{\mathbf{m}}$. However, for $\mathbf{m} = \mathbf{0}$ it requires that $C_{\mathbf{m}} = 0$, whereas from equation (3.5.2.9) $C_{\mathbf{m}} = 1/V$. To avoid this contradiction we modify ρ^{P} by subtracting $1/V$ (uniform neutralizing plasma). That is, ρ^{P} is now defined by

$$\rho^{\mathrm{P}}(\mathbf{r}) = \sum_{\mathbf{n}}\rho(\mathbf{r}+\mathbf{n}) - (1/V). \qquad (3.5.2.12)$$

Note that ρ^{P} is still periodic. Also note that for $\mathbf{m}\neq\mathbf{0}$ $C_{\mathbf{m}}$ is still given by equation (3.5.2.9), whereas now $C_{\mathbf{0}} = 0$. Although we have avoided the contradiction, the value of $B_{\mathbf{0}}$ is still not determined. Some authors define it by considering $\mathbf{m}$ as a real variable and taking the limit as $\mathbf{m}\to\mathbf{0}$ (this depends on which direction the origin is approached from), but most treatments take it to be 0 (however, see Section 3.5.2.3). In the latter case the electrostatic potential Φ^{P} due to ρ^{P} is given by

$$\Phi^{\mathrm{P}}(\mathbf{r}) = \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}\exp(2\pi i\mathbf{m}\cdot\mathbf{r}). \qquad (3.5.2.13)$$

Now as before let $q_1, q_2, \ldots, q_N$ be point charges located at positions $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ within the unit cell U. From the discussion in Section 3.5.2.1 we assume the unit cell is *neutral*, *i.e.* $q_1 + \ldots + q_N = 0$. Let $\mathbf{r}$ be a point in U with $\mathbf{r}\neq\mathbf{r}_i$, $i = 1, \ldots, N$. We want to obtain the electrostatic potential Φ due to the point charges in U together with their periodic images in the remainder of the crystal C. That is,

$$\Phi_{\mathrm{lattice}}(\mathbf{r}) = \sum_{i=1}^{N}\sum_{\mathbf{n}}\frac{q_i}{|\mathbf{r}_i+\mathbf{n}-\mathbf{r}|}.$$

We next supplement the potential due to the point charge q_i at $\mathbf{r}_i+\mathbf{n}$ by the potential due to an equally charged Gaussian counterion, that is $(-q_i)$ times a density of the form ρ in equation (3.5.2.6), centred at $\mathbf{r}_i+\mathbf{n}$. Using equation (3.5.2.7) the resulting *direct sum* potential is given by

$$\Phi_{\mathrm{dir}}(\mathbf{r}) = \sum_{i=1}^{N}\sum_{\mathbf{n}}\frac{q_i\,\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}_i+\mathbf{n}-\mathbf{r}|)}{|\mathbf{r}_i+\mathbf{n}-\mathbf{r}|},$$

where $\mathrm{erfc}(x) = 1-\mathrm{erf}(x)$. Note that since $\mathrm{erfc}(x)$ decays exponentially fast to zero as $x\to\infty$, the sum in the above equation converges rapidly and absolutely.

However, having added the potential due to counterions we now need to cancel it. That is, we must consider the *reciprocal* electrostatic potential due to $(+q_i)$ times a density of the form ρ, centred at $\mathbf{r}_i+\mathbf{n}$, for all direct-lattice vectors $\mathbf{n}$. Since the unit cell is neutral, the sum of these *reciprocal* potentials over $i = 1, \ldots, N$ and all $\mathbf{n}$ is the same as the sum over $i = 1, \ldots, N$ of q_i times the potential due to ρ^{P} centred at $\mathbf{r}_i$ where ρ^{P} is given by equation (3.5.2.12). Note that the extra term involving $1/V$ drops out by neutrality. [In the case of non-neutrality, such as discussion of the free energy of charging an ion (Hummer *et al.*, 1996), the potential due to the uniform compensating plasma must be included (see below). Hummer has derived the potential due to a uniformly charged unit cell in the special case of a cubic unit cell (Hummer, 1996).] Thus, using equation (3.5.2.13) we arrive at the *reciprocal sum* potential:

$$
\begin{aligned}
\Phi_{\mathrm{rec}}(\mathbf{r}) &= \sum_{i=1}^{N} q_i\Phi^{\mathrm{P}}(\mathbf{r}-\mathbf{r}_i) \\
&= \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2} \\
&\quad\times\sum_{i=1}^{N} q_i\exp[2\pi i\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_i)] \\
&= \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}\exp(2\pi i\mathbf{m}\cdot\mathbf{r})S(-\mathbf{m}),
\end{aligned}
$$

where the structure factor $S(\mathbf{m})$ is given by

$$S(\mathbf{m}) = \sum_{i}^{N} q_i\exp(2\pi i\mathbf{m}\cdot\mathbf{r}_i). \qquad (3.5.2.14)$$

We have now expressed $\Phi_{\mathrm{lattice}}(\mathbf{r})$ as the sum of $\Phi_{\mathrm{dir}}(\mathbf{r})$ plus $\Phi_{\mathrm{rec}}(\mathbf{r})$ (with the caveat that in the process some infinite series have been casually rearranged and that the contribution of the reciprocal vector $\mathbf{m} = \mathbf{0}$ has been ignored). To get a valid electrostatic potential $\Phi_{\mathrm{lattice}}(\mathbf{r})$ at the position of a charge q_i in U we must remove the Coulomb potential due to q_i and then take the limit as $\mathbf{r}\to\mathbf{r}_i$. Noting that $1 = \mathrm{erf}(x) + \mathrm{erfc}(x)$ and that $\lim_{x\to 0}\mathrm{erf}(x)/x = 2/\sqrt{\pi}$, we have

$$\begin{aligned}\Phi_{\rm lattice}(\mathbf{r}_i) &= \lim_{\mathbf{r}\to\mathbf{r}_i}\left[\Phi_{\rm lattice}(\mathbf{r}) - \frac{q_i}{|\mathbf{r}-\mathbf{r}_i|}\right] \\ &= \sum_{j\neq i}\frac{q_j\ \mathrm{erfc}(\alpha^{1/2}|\mathbf{r}_j-\mathbf{r}_i|)}{|\mathbf{r}_j-\mathbf{r}_i|} \\ &\quad + \sum_{j=1}^{N}\sum_{\mathbf{n}\neq\mathbf{0}}\frac{q_j\ \mathrm{erfc}(\alpha^{1/2}|\mathbf{r}_j+\mathbf{n}-\mathbf{r}_i|)}{|\mathbf{r}_j+\mathbf{n}-\mathbf{r}_i|} \\ &\quad + \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}\exp(2\pi i\mathbf{m}\cdot\mathbf{r}_i)S(-\mathbf{m}) \\ &\quad - 2q_i\Big(\frac{\alpha}{\pi}\Big)^{1/2}.\end{aligned}$$

Thus we can write the electrostatic energy of U from equation (3.5.2.3) as

$$\begin{aligned}E_{\rm Ewald}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2}\sum_{i=1}^{N} q_i\,\Phi_{\rm lattice}(\mathbf{r}_i) \\ &= \tfrac{1}{2}{\sum_{\mathbf{n}}}'\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_iq_j\ \mathrm{erfc}(\alpha^{1/2}|\mathbf{r}_j+\mathbf{n}-\mathbf{r}_i|)}{|\mathbf{r}_j+\mathbf{n}-\mathbf{r}_i|} \\ &\quad + \frac{1}{2\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}S(\mathbf{m})S(-\mathbf{m}) \\ &\quad - \Big(\frac{\alpha}{\pi}\Big)^{1/2}\sum_{i=1}^{N}q_i^2, \qquad (3.5.2.15)\end{aligned}$$

where as above the prime indicates that terms with $i=j$ and $\mathbf{n}=\mathbf{0}$ are omitted.

3.5.2.3. Ewald sum: more complete derivation

The derivation in Section 3.5.2.2 is incomplete. In addition to the caveats concerning the expression for $\Phi_{\rm lattice}(\mathbf{r})$ in terms of $\Phi_{\rm dir}(\mathbf{r})$ plus $\Phi_{\rm rec}(\mathbf{r})$, there is no mention of a quadratic term involving the unit-cell dipole moment $\mathbf{M}$. In this section we provide a more complete derivation of the Ewald sum, which will in addition explain the conditional convergence discussed previously. We follow the developments in Smith (1981, 1994).

The derivation depends on the following identity:

$$\begin{aligned}\frac{1}{|\mathbf{r}|} &= \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}|)}{|\mathbf{r}|} + \frac{1}{\pi}\int_{\Re^3}\frac{\exp(-\pi^2\mathbf{u}^2/\alpha)}{\mathbf{u}^2}\exp(-2\pi i\mathbf{u}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{u} \\ &= \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}|)}{|\mathbf{r}|} + \frac{1}{\pi}\sum_{\mathbf{m}}\int_{U^*}\frac{\exp(-\pi^2|\mathbf{v}+\mathbf{m}|^2/\alpha)}{|\mathbf{v}+\mathbf{m}|^2} \\ &\quad\times\exp[-2\pi i(\mathbf{v}+\mathbf{m})\cdot\mathbf{r}]\,\mathrm{d}^3\mathbf{v}, \qquad (3.5.2.16)\end{aligned}$$

where U^* is the reciprocal unit cell, *i.e.* the set of points $\mathbf{v}=t_1a_1^*+t_2a_2^*+t_3a_3^*$ with $-\frac{1}{2}\le t_\alpha\le\frac{1}{2}$, $\alpha=1,2,3$. Actually we derive (following Essmann *et al.*, 1995) a generalization of identity (3.5.2.16), valid for general inverse powers $1/|\mathbf{r}|^p$, $p>0$. For this first note that for $u>0$

$$\begin{aligned}\Gamma(u) &= \int_0^\infty t^{u-1}\exp(-t)\,\mathrm{d}t \\ &= \lambda^u\int_0^\infty t^{u-1}\exp(-\lambda t)\,\mathrm{d}t, \qquad (3.5.2.17)\end{aligned}$$

where $\Gamma(u)$ is the Euler gamma function. Next note that for $a>0$

$$\begin{aligned}\exp(-a\mathbf{r}^2) &= (\pi/a)^{3/2}\int_{\Re^3}\exp(-\pi^2\mathbf{u}^2/a) \\ &\quad\times\exp(-2\pi i\mathbf{u}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{u}, \qquad (3.5.2.18)\end{aligned}$$

which is the Fourier transform of the Gaussian, as derived above within equation (3.5.2.9). In equation (3.5.2.17) substitute $\lambda=|\mathbf{r}|^2$ and $u=p/2$ to get, for $\alpha>0$,

$$\begin{aligned}\frac{\Gamma(p/2)}{|\mathbf{r}|^p} &= \int_0^\alpha t^{p/2-1}\exp(-\mathbf{r}^2t)\,\mathrm{d}t + \int_\alpha^\infty t^{p/2-1}\exp(-\mathbf{r}^2t)\,\mathrm{d}t \\ &= \mathrm{I}_p + \mathrm{II}_p.\end{aligned}$$

In II_p we substitute $\mathbf{r}^2t=s^2$, whereas in I_p we use equation (3.5.2.18), change the order of integration, and substitute $\pi^2\mathbf{u}^2/t=s^2$ to obtain

$$\begin{aligned}\frac{1}{|\mathbf{r}|^p} &= \frac{1}{\Gamma(p/2)}\frac{1}{|\mathbf{r}|^p}\int_{\alpha^{1/2}|\mathbf{r}|}^\infty 2s^{p-1}\exp(-s^2)\,\mathrm{d}s \\ &\quad + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)}\int_{\Re^3}\mathrm{d}^3\mathbf{u}\exp(-2\pi i\mathbf{u}\cdot\mathbf{r})\Big(\frac{\pi|\mathbf{u}|}{\alpha^{1/2}}\Big)^{p-3} \\ &\quad\times\int_{\pi|\mathbf{u}|/\alpha^{1/2}}^\infty 2s^{2-p}\exp(-s^2)\,\mathrm{d}s \\ &= \frac{f_p(\alpha^{1/2}|\mathbf{r}|)}{|\mathbf{r}|^p} \\ &\quad + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)}\int_{\Re^3}\exp(-2\pi i\mathbf{u}\cdot\mathbf{r})g_p(\pi|\mathbf{u}|/\alpha^{1/2})\,\mathrm{d}^3\mathbf{u} \\ &= \frac{f_p(\alpha^{1/2}|\mathbf{r}|)}{|\mathbf{r}|^p} \\ &\quad + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)}\sum_{\mathbf{m}}\int_{U^*}g_p\Big(\frac{\pi|\mathbf{v}+\mathbf{m}|}{\alpha^{1/2}}\Big) \\ &\quad\times\exp[-2\pi i(\mathbf{v}+\mathbf{m})\cdot\mathbf{r}]\,\mathrm{d}^3\mathbf{v}, \qquad (3.5.2.19)\end{aligned}$$

where

$$f_p(x) = \frac{1}{\Gamma(p/2)}\int_x^\infty 2s^{p-1}\exp(-s^2)\,\mathrm{d}s \qquad (3.5.2.20)$$

and

$$g_p(x) = x^{p-3}\int_x^\infty 2s^{2-p}\exp(-s^2)\,\mathrm{d}s. \qquad (3.5.2.21)$$

Specializing to the case $p=1$ we see that $f_p(x)=\mathrm{erfc}(x)$ and that $g_p(x)=\exp(-x^2)/x^2$, which proves identity (3.5.2.16). Another important case is $p=6$, for which $f_p(x)=(1+x^2+x^4/2)\exp(-x^2)$ and $g_p(x)=(1-2x^2)\exp(-x^2)/3+2x^3\pi^{1/2}\mathrm{erfc}(x)$.

Note that from equation (3.5.2.17), substituting $t=s^2$, $\lambda=\mathbf{r}^2$ and $u=p/2$ we have

$$\frac{1}{|\mathbf{r}|^p} = \frac{2}{\Gamma(p/2)} \int_0^\infty s^{p-1} \exp(-r^2 s^2)\, \mathrm{d}s.$$

Also, substituting $s = |\mathbf{r}|t$ into equation (3.5.2.20) we have

$$\frac{f_p(\alpha^{1/2}|\mathbf{r}|)}{|\mathbf{r}|^p} = \frac{2}{\Gamma(p/2)} \int_{\alpha^{1/2}}^\infty t^{p-1} \exp(-r^2 t^2)\, \mathrm{d}t$$

and so

$$\begin{aligned}\lim_{\mathbf{r}\to\mathbf{0}} &\frac{1}{|\mathbf{r}|^p}[1 - f_p(\alpha^{1/2}|\mathbf{r}|)] \\ &= \lim_{\mathbf{r}\to\mathbf{0}} \frac{2}{\Gamma(p/2)} \int_0^{\alpha^{1/2}} t^{p-1} \exp(-r^2 t^2)\, \mathrm{d}t \\ &= \frac{2\alpha^{p/2}}{p\Gamma(p/2)}. \end{aligned} \tag{3.5.2.22}$$

Note that $g_p(x)$ is a smooth function of x for $x \neq 0$, and that its derivatives are uniformly bounded for x bounded away from 0. To show this, substitute $s = xt$ in equation (3.5.2.21) to write

$$g_p(x) = 2 \int_1^\infty t^{2-p} \exp(-x^2 t^2)\, \mathrm{d}t.$$

In the case that $p = 6$, $g_p(x)$ is smooth with uniformly bounded derivatives for all x, but this is unfortunately not true for all p. However, from the above expression we can show that for $p > 3$, $g_p(x)$ is bounded and continuous as $x \to 0$, whereas for $1 \leq p \leq 3$ $g_p(x) \to \infty$ as $x \to 0$. Thus for $p > 3$ we see that for any $x > 0$, $g_p(x) < g_p(0) = 2/(p-3)$. However for $1 \leq p \leq 3$, for any $R > 1$ if $x > 0$ sufficiently small $\exp(-x^2R^2) > \frac{1}{2}$; thus $g_p(x) > 2\int_1^R t^{2-p} \exp(-x^2t^2)\, \mathrm{d}t > \int_1^R t^{2-p}\, \mathrm{d}t \geq \ln(R)$ and thus $g_p(x)$ is unbounded as $x \to 0$.

Despite this latter result, the integral of $g_p(\mathbf{r})$ over $\Re^3$ is bounded for $p \geq 1$. This result is important below for the validity of rearranging the order of summation or taking limits inside infinite sums involving g_p. To prove boundedness, use the above expression for g_p and integrate by parts to write

$$\begin{aligned}\int_{\Re^3} g_p(\mathbf{r})\, \mathrm{d}^3\mathbf{r} &= 4\pi \int_0^\infty x^2 g_p(x)\, \mathrm{d}x \\ &= 4\pi \int_0^\infty \mathrm{d}x \int_1^\infty t^{1-p}(2x^2 t) \exp(-x^2t^2)\, \mathrm{d}t \\ &= 4\pi \Bigg[\int_0^\infty \exp(-x^2)\, \mathrm{d}x \\ &\quad + (1-p) \int_1^\infty t^{-p}\, \mathrm{d}t \int_0^\infty \exp(-x^2t^2)\, \mathrm{d}x \Bigg] \\ &= 4\pi \Bigg[\int_0^\infty \exp(-x^2)\, \mathrm{d}x \\ &\quad + (1-p) \int_1^\infty t^{-(p+1)}\, \mathrm{d}t \int_0^\infty \exp(-s^2)\, \mathrm{d}s \Bigg] \\ &= \frac{2\pi^{3/2}}{p}. \end{aligned}$$

3.5.2.3.1. The case of $p > 3$

We first discuss the lattice sum for the case $p > 3$. In this case the lattice sum converges absolutely, since the tail of this sum can be bounded from above by a constant times the integral

$$\int_{|\mathbf{r}|>K} \frac{1}{|\mathbf{r}|^p}\, \mathrm{d}^3\mathbf{r} = 4\pi \int_K^\infty r^{(2-p)}\, \mathrm{d}r = \frac{K^{(3-p)}}{p-3} \tag{3.5.2.23}$$

for some $K > 0$. Although the sum can thus be evaluated directly, it can be significantly accelerated using the same Ewald direct- and reciprocal-space decomposition. For $\mathbf{r} \neq \mathbf{0}$, using equation (3.5.2.19) we have

$$\begin{aligned}\sum_{\mathbf{n}} \frac{1}{|\mathbf{r}+\mathbf{n}|^p} = \sum_{\mathbf{n}} \frac{f_p(\alpha^{1/2}|\mathbf{r}+\mathbf{n}|)}{|\mathbf{r}+\mathbf{n}|^p} + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)} \\ \times \sum_{\mathbf{n}} \sum_{\mathbf{m}} \int_{U^*} h_{p,\mathbf{m},\mathbf{r}}(\mathbf{v}) \exp(-2\pi i \mathbf{v}\cdot\mathbf{n})\, \mathrm{d}^3\mathbf{v}, \end{aligned}$$

where $h_{p,\mathbf{m},\mathbf{r}}(\mathbf{v}) = g_p(\pi|\mathbf{v}+\mathbf{m}|/\alpha^{1/2}) \exp[-2\pi i(\mathbf{v}+\mathbf{m})\cdot\mathbf{r}]$. Since $f_p(\alpha^{1/2}|\mathbf{r}+\mathbf{n}|)$ is bounded, the first of the above sums converges absolutely and we turn to the second. We recognize the integral $\int_{U^*} h_{p,\mathbf{m},\mathbf{r}}(\mathbf{v}) \exp(-2\pi i \mathbf{v}\cdot\mathbf{n})\, \mathrm{d}^3\mathbf{v}$ as $(1/V)$ times $\hat{h}_{p,\mathbf{m},\mathbf{r}}(\mathbf{n})$, the $\mathbf{n}$th Fourier coefficient of $h_{p,\mathbf{m},\mathbf{r}}$, where V is the volume of the unit cell U ($1/V$ is the volume of U^*). From the above results for g_p it is clear that for $\mathbf{m} \neq \mathbf{0}$, $h_{p,\mathbf{m},\mathbf{r}}$ are smooth and have uniformly (in $\mathbf{v}$, $\mathbf{r}$ and $\mathbf{m} \neq \mathbf{0}$) bounded derivatives. Thus the sum of the Fourier coefficients can be shown to converge absolutely and uniformly in $\mathbf{m}$, which guarantees uniform (over $\mathbf{m} \neq \mathbf{0}$) convergence of the sum of these integrals to $(1/V)h_{p,\mathbf{m},\mathbf{r}}(\mathbf{0})$, which in turn equals $(1/V)g_p(\pi|\mathbf{m}|/\alpha^{1/2}) \exp(-2\pi i \mathbf{m}\cdot\mathbf{r})$.

The case of $\mathbf{m} = \mathbf{0}$ is not as straightforward, since the uniform boundedness of derivatives of $h_{p,\mathbf{m},\mathbf{r}}$ does not generally hold in this case. One way of handling it is as follows. Since everything else in the above equation converges as $|\mathbf{n}| \to \infty$, the sum of the Fourier coefficients for $h_{p,\mathbf{m}=\mathbf{0},\mathbf{r}}$ converges to something. That is, if

$$S(n_1, n_2, n_3) = \sum_{k_1=-n_1}^{n_1} \sum_{k_2=-n_2}^{n_2} \sum_{k_3=-n_3}^{n_3} \hat{h}_{p,\mathbf{0},\mathbf{r}}(k_1, k_2, k_3)$$

for $n_\alpha > 0$, $\alpha = 1, 2, 3$, then for some s, $S(n_1, n_2, n_3) \to s$ as $\max(n_1, n_2, n_3) \to \infty$. If

$$\sigma(n) = [1/(n+1)]^3 \sum_{n_1=0}^{n} \sum_{n_2=0}^{n} \sum_{n_3=0}^{n} S(n_1, n_2, n_3),$$

then it is straightforward to show that $\sigma(n)$ also converges to s as $n \to \infty$ [see, for example, Chapter 1 of Körner (1988) for the one-dimensional proof]. On the other hand, due to the boundedness and continuity of h_p, $\sigma(n) \to h_{p,\mathbf{m},\mathbf{r}}(\mathbf{0})$ as $n \to \infty$. [This involves the three-dimensional version of the Fejér kernel, which converges to the three-dimensional Dirac delta function. See, for example, Chapter 2 of Körner (1988) for the one-dimensional proof.] Thus $S(n_1, n_2, n_3) \to h_{p,\mathbf{m},\mathbf{r}}(\mathbf{0})$ as $\max(n_1, n_2, n_3) \to \infty$. Hence the above sum of integrals with $\mathbf{m} = \mathbf{0}$ converges to $(1/V)h_{p,\mathbf{0},\mathbf{r}}(\mathbf{0})$.

Thus from the above arguments we have

$$\sum_{\mathbf{n}} \frac{1}{|\mathbf{r}+\mathbf{n}|^p} = \sum_{\mathbf{n}} \frac{f_p(\alpha^{1/2}|\mathbf{r}+\mathbf{n}|)}{|\mathbf{r}+\mathbf{n}|^p} + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)} \frac{1}{V} \sum_{\mathbf{m}} g_p(\pi|\mathbf{m}|/\alpha^{1/2}) \times \exp(-2\pi i \mathbf{m}\cdot\mathbf{r}). \quad (3.5.2.24)$$

Using the limiting behaviour given in (3.5.2.22) we can extend this result to the case $\mathbf{r} = \mathbf{0}$. By removing the singularity when $\mathbf{n} = \mathbf{0}$ we get

$$\begin{aligned}\sum_{\mathbf{n}\neq\mathbf{0}} \frac{1}{|\mathbf{n}|^p} &= \lim_{\mathbf{r}\to\mathbf{0}} \sum_{\mathbf{n}\neq\mathbf{0}} \frac{1}{|\mathbf{r}+\mathbf{n}|^p} \\ &= \lim_{\mathbf{r}\to\mathbf{0}} \left(\sum_{\mathbf{n}} \frac{1}{|\mathbf{r}+\mathbf{n}|^p} - \frac{1}{|\mathbf{r}|^p} \right) \\ &= \sum_{\mathbf{n}\neq\mathbf{0}} \frac{f_p(\alpha^{1/2}|\mathbf{n}|)}{|\mathbf{n}|^p} - \frac{2\alpha^{p/2}}{p\Gamma(p/2)} + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)} \\ &\quad \times \frac{1}{V} \sum_{\mathbf{m}} g_p(\pi|\mathbf{m}|/\alpha^{1/2}). \end{aligned} \quad (3.5.2.25)$$

Now assume particles i and j in the unit cell U interact by the potential $\phi_{ij} = C(i,j)/|\mathbf{r}_j - \mathbf{r}_i|^p$. Using equations (3.5.2.24) and (3.5.2.25) (the latter for particles interacting with their own images) the total energy for the unit cell U consisting of N particles interacting *via* the potential φ with each other and all of the image cells is given by

$$\begin{aligned} E(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{C(i,j)}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{n}|^p} \\ &= \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{C(i,j) f_p(\alpha^{1/2}|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|)}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|^p} \\ &\quad + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)} \frac{1}{2V} \sum_{\mathbf{m}} g_p(\pi|\mathbf{m}|/\alpha^{1/2}) \\ &\quad \times \sum_{i=1}^{N}\sum_{j=1}^{N} C(i,j) \exp[-2\pi i \mathbf{m}\cdot(\mathbf{r}_j - \mathbf{r}_i)] \\ &\quad - \frac{\alpha^{p/2}}{p\Gamma(p/2)} \sum_{i=1}^{N} C(i,i), \end{aligned}$$

where as before the prime on the first summation means that terms where $i = j$ and $\mathbf{n} = \mathbf{0}$ are omitted.

In the case when $C(i,j)$ factorizes, *i.e.* $C(i,j) = C(i)C(j)$, the above lattice sum can be further simplified and accelerated using the structure factor $S(\mathbf{m})$ given by

$$S(\mathbf{m}) = \sum_{j=1}^{N} C(j) \exp(2\pi i \mathbf{m}\cdot\mathbf{r}_i). \quad (3.5.2.26)$$

In this case,

$$\begin{aligned} E(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{C(i)C(j) f_p(\alpha^{1/2}|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|)}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|^p} \\ &\quad + \frac{\pi^{3/2}\alpha^{(p-3)/2}}{\Gamma(p/2)} \frac{1}{2V} \sum_{\mathbf{m}} g_p(\pi|\mathbf{m}|/\alpha^{1/2}) S(\mathbf{m}) S(-\mathbf{m}) \\ &\quad - \frac{\alpha^{p/2}}{p\Gamma(p/2)} \sum_{i=1}^{N} C(i)^2. \end{aligned}$$

Note that as expected, due to the absolute convergence of the lattice sums, there is no shape-dependent correction term to the Ewald sum. This is connected with the fact that $g_p(x)$ is bounded and continuous as $x \to 0$. Next we see what additional assumptions are necessary when this is no longer true, and what form the correction term takes. We cover the most interesting case, namely the Coulombic case $p = 1$.

3.5.2.3.2. The case p = 1

In the case $1 \le p \le 3$, the lattice sum does not converge absolutely, since the integral in equation (3.5.2.23) diverges. Thus extra conditions are needed and the order of summation must be specified. We restrict discussion to the case $p = 1$, that is to a crystal of point charges interacting *via* Coulomb's law. We consider a large but finite crystal, taking the limit as it grows larger while maintaining a specific shape. To that end, let P be a closed, bounded region in $\Re^3$ centred on the origin, such as a cube, sphere or ellipsoid. For a positive integer K, let $\Omega(P,K)$ denote the set of lattice vectors such that $|\mathbf{n}|/K$ is in P, that is $\Omega(P,K) = \{\mathbf{n} : |\mathbf{n}|/K \in P\}$. Charges in the central unit cell $\mathbf{U}_0$ interact with each other and with image charges in $\mathbf{U}_\mathbf{n}$ for $\mathbf{n} \in \Omega(P,K)$. The electrostatic energy of the central cell in this finite crystal is given by

$$E_{P,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) = \tfrac{1}{2} {\sum_{\mathbf{n}\in\Omega(P,K)}}' \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{n}|} \quad (3.5.2.27)$$

and we want the limit of $E_{P,K}$ as $K \to \infty$.

Recalling that $g_p(x) = \exp(-x^2)/x^2$ for $p = 1$, if we define $h_{\mathbf{m},\mathbf{r}}(\mathbf{v})$ by

$$h_{\mathbf{m},\mathbf{r}}(\mathbf{v}) = \frac{\exp[-\pi^2(\mathbf{m}+\mathbf{v})^2/\alpha]}{|\mathbf{m}+\mathbf{v}|^2} \exp[-2\pi i(\mathbf{v}+\mathbf{m})\cdot\mathbf{r})],$$

then for $\mathbf{m} \neq \mathbf{0}$, $h_{\mathbf{m},\mathbf{r}}(\mathbf{v})$ is bounded and smooth with uniformly (in $\mathbf{m} \neq \mathbf{0}$) bounded derivatives. For $\mathbf{m} = \mathbf{0}$ we have $h_{\mathbf{m},\mathbf{r}}(\mathbf{v}) = [1 - 2\pi i(\mathbf{v}\cdot\mathbf{r}) - 2\pi^2(\mathbf{v}\cdot\mathbf{r})^2]/\mathbf{v}^2 - \pi^2/\alpha + \theta(\mathbf{v})$, where $\theta(\mathbf{v})$ is smooth with bounded derivatives, and with $\theta(\mathbf{0}) = 0$. Then, using the same arguments that lead to equation (3.5.2.24), we can write for $\mathbf{r} \neq \mathbf{0}$

$$\begin{aligned} \sum_{\mathbf{n}\in\Omega(P,K)} \frac{1}{|\mathbf{r}+\mathbf{n}|} &= \sum_{\mathbf{n}} \frac{\operatorname{erfc}(\alpha^{1/2}|\mathbf{r}+\mathbf{n}|)}{|\mathbf{r}+\mathbf{n}|} \\ &\quad + \frac{1}{\pi V} \sum_{\mathbf{m}\neq\mathbf{0}} \frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2} \exp(-2\pi i \mathbf{m}\cdot\mathbf{r}) \\ &\quad - \frac{\pi}{\alpha V} + \frac{1}{\pi} H_{P,K}(\mathbf{r}) + \varepsilon(K) \\ &= \varphi_{\text{Ewald}}(\mathbf{r}) + \frac{1}{\pi} H_{P,K}(\mathbf{r}) + \varepsilon(K) \end{aligned} \quad (3.5.2.28)$$

where here and below $\varepsilon(K)$ denotes a quantity (the remainder in the equation) that converges to zero as $K \to \infty$ and

$$H_{P,K}(\mathbf{r}) = \sum_{\mathbf{n}\in\Omega(P,K)} \int_{U^*} \frac{1 - 2\pi i(\mathbf{v}\cdot\mathbf{r}) - 2\pi^2(\mathbf{v}\cdot\mathbf{r})^2}{\mathbf{v}^2} \times \exp(-2\pi i\mathbf{v}\cdot\mathbf{n})\,\mathrm{d}^3\mathbf{v}.$$

The Ewald potential $\varphi_{\mathrm{Ewald}}(\mathbf{r})$, given for $\mathbf{r}\neq\mathbf{0}$ by

$$\varphi_{\mathrm{Ewald}}(\mathbf{r}) = \sum_{\mathbf{n}} \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}+\mathbf{n}|)}{|\mathbf{r}+\mathbf{n}|} - \frac{\pi}{\alpha V} + \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}} \frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}\exp(-2\pi i\mathbf{m}\cdot\mathbf{r}),$$

has a couple of noteworthy properties. First, note that since the left-hand side of equation (3.5.2.28) is independent of α, as is $H_{P,K}(\mathbf{r})$, $\varphi_{\mathrm{Ewald}}(\mathbf{r})$ is itself independent of α. Secondly, the average of $\varphi_{\mathrm{Ewald}}(\mathbf{r})$ over the unit cell U is zero. To establish this latter property, first note that for $\mathbf{m}\neq\mathbf{0}$, $\int_U \exp(2\pi\mathbf{m}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{r} = 0$. Thus the reciprocal sum in the above equation integrates to zero over U. Next, the integral of $-\pi/(\alpha V)$ over U is clearly $-\pi/\alpha$. Finally, note that

$$\begin{aligned}
&\sum_{\mathbf{n}}\int_U \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}+\mathbf{n}|)}{|\mathbf{r}+\mathbf{n}|}\,\mathrm{d}^3\mathbf{r}\\
&\quad= \int_{\Re^3} \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}|)}{|\mathbf{r}|}\,\mathrm{d}^3\mathbf{r}\\
&\quad= 8\pi^{1/2}\int_{r=0}^{\infty} r\,\mathrm{d}r \int_{\alpha^{1/2}r}^{\infty} \exp(-t^2)\,\mathrm{d}t\\
&\quad= 4\pi^{1/2}\int_0^{\infty} \exp(-t^2)\,\mathrm{d}t \int_0^{t/\alpha^{1/2}} 2r\,\mathrm{d}r\\
&\quad= \frac{4\pi^{1/2}}{\alpha}\int_0^{\infty} t^2\exp(-t^2)\,\mathrm{d}t\\
&\quad= \frac{\pi}{\alpha},
\end{aligned}$$

completing the proof. The self or Wigner potential $q\zeta$ with ζ given by

$$\begin{aligned}
\zeta &= \lim_{|\mathbf{r}|\to 0}\left[\varphi_{\mathrm{Ewald}}(\mathbf{r}) - \frac{1}{|\mathbf{r}|}\right]\\
&= -2\left(\frac{\alpha}{\pi}\right)^{1/2} + \sum_{\mathbf{n}\neq\mathbf{0}} \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{n}|)}{|\mathbf{n}|}\\
&\quad - \frac{\pi}{\alpha V} + \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}} \frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}
\end{aligned}$$

is the potential of a charge q in a periodic array of images of itself together with a uniform neutralizing plasma throughout the crystal. For cubic unit cells with side L, $\zeta \simeq -2.837297/L$ (Nijboer & Ruijgrok, 1988). Owing to the above properties, the Ewald potential provides a consistent approach to the treatment of non-neutral unit cells, such as when calculating ionic charging free energies (Hummer *et al.*, 1996; Darden *et al.*, 1999).

If we define the structure factor $S(\mathbf{m})$ by

$$S(\mathbf{m}) = \sum_{j=1}^{N} q_j \exp(2\pi i\mathbf{m}\cdot\mathbf{r}_j)$$

and the unit-cell dipole moment $\mathbf{D}$ by

$$\mathbf{D} = \sum_{j=1}^{N} q_j\mathbf{r}_j,$$

then assuming the unit cell is neutral we can write the energy $E_{P,K}$ as

$$\begin{aligned}
E_{P,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{i=1}^{N}\sum_{j=1}^{N} q_iq_j \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}_j-\mathbf{r}_i+\mathbf{n}|)}{|\mathbf{r}_j-\mathbf{r}_i+\mathbf{n}|}\\
&\quad + \frac{1}{2\pi V}\sum_{\mathbf{m}\neq\mathbf{0}} \frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2} S(\mathbf{m})S(-\mathbf{m})\\
&\quad + 2\pi\sum_{\mathbf{n}\in\Omega(P,K)}\int_{U^*} \frac{(\mathbf{v}\cdot\mathbf{D})^2}{\mathbf{v}^2}\exp(-2\pi i\mathbf{v}\cdot\mathbf{n})\,\mathrm{d}^3\mathbf{v}\\
&\quad - \left(\frac{\alpha}{\pi}\right)^{1/2}\sum_{j=1}^{N} q_j^2 + \varepsilon(K). \qquad (3.5.2.29)
\end{aligned}$$

Thus, recalling equation (3.5.2.15), we can write

$$E_{P,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) = E_{\mathrm{Ewald}}(\underline{\mathbf{r}}^{\mathbf{N}}) + J(\mathbf{D},P,K) + \varepsilon(K), \qquad (3.5.2.30)$$

where

$$J(\mathbf{D},P,K) = 2\pi\sum_{\mathbf{n}\in\Omega(P,K)}\int_{U^*}\frac{(\mathbf{v}\cdot\mathbf{D})^2}{\mathbf{v}^2}\exp(-2\pi i\mathbf{v}\cdot\mathbf{n})\,\mathrm{d}^3\mathbf{v}.$$

Note that if $\mathbf{D}=\mathbf{0}$, $J(\mathbf{D},P,K)=0$ and the Coulomb lattice sum converges absolutely, *i.e.* the result is the same as the Ewald sum regardless of the shape P, *i.e.* order of convergence. Expressions for $J(\mathbf{D},P) = \lim_{K\to\infty} J(\mathbf{D},P,K)$ have been derived in some cases by Smith (1981). In particular, if P is the unit sphere $J(\mathbf{D},P) = 2\pi\mathbf{D}^2/(3V)$. Following Smith's approach we show this by first deriving an integral representation for $J(\mathbf{D},P)$ and then evaluating it for the case P is the unit sphere.

First note that for any continuous function f, $\sum_{\mathbf{n}\in\Omega(P,K)} f(\mathbf{n}/K)V/K^3$ converges to $\int_P f(\rho)\,\mathrm{d}^3\rho$ as $K\to\infty$. Then, in the above integral for $J(\mathbf{D},P,K)$ we can substitute $\mathbf{w} = K\mathbf{v}$ to get

$$\begin{aligned}
J(\mathbf{D},P,K) &= \frac{2\pi}{V}\int_{K\cdot U^*}\frac{(\mathbf{w}\cdot\mathbf{D})^2}{\mathbf{w}^2}\,\mathrm{d}^3\mathbf{w}\\
&\quad\times\sum_{\mathbf{n}\in\Omega(P,K)}\exp\left(-2\pi i\mathbf{w}\cdot\frac{\mathbf{n}}{K}\right)\frac{V}{K^3}\\
&= \frac{2\pi}{V}\int_{K\cdot U^*}\frac{(\mathbf{w}\cdot\mathbf{D})^2}{\mathbf{w}^2}\,\mathrm{d}^3\mathbf{w}\\
&\quad\times\int_P\exp(-2\pi i\mathbf{w}\cdot\rho)\,\mathrm{d}^3\rho + \varepsilon(K)
\end{aligned}$$

and thus

$$J(\mathbf{D}, P) = \frac{2\pi}{V} \int_{\Re^3} \frac{(\mathbf{w}\cdot\mathbf{D})^2}{\mathbf{w}^2}\, \mathrm{d}^3\mathbf{w} \int_P \exp(-2\pi i \mathbf{w}\cdot\rho)\, \mathrm{d}^3\rho.$$

Specializing to the case $P = S$, the unit sphere, we see by using spherical coordinates for ρ with the pole parallel to $\mathbf{w}$ that $\int_S \exp(-2\pi i\mathbf{w}\cdot\rho)\,\mathrm{d}^3\rho$ is a function only of the norm of $\mathbf{w}$, that is $\int_S \exp(-2\pi i\mathbf{w}\cdot\rho)\,\mathrm{d}^3\rho = g(|\mathbf{w}|)$. Then, again using spherical coordinates, this time for $\mathbf{w}$ with the pole parallel to $\mathbf{D}$, and noting that $\int_{\theta=0}^{\pi} \cos^2(\theta)\sin(\theta)\,\mathrm{d}\theta = \frac{1}{3}\int_{\theta=0}^{\pi}\sin(\theta)\,\mathrm{d}\theta$, we have

$$\begin{aligned}
J(\mathbf{D}, S) &= \frac{2\pi}{V} \int_{\Re^3} \frac{(\mathbf{w}\cdot\mathbf{D})^2}{\mathbf{w}^2} g(|\mathbf{w}|)\, \mathrm{d}^3\mathbf{w} \\
&= \frac{2\pi\mathbf{D}^2}{V} \int_{w=0}^{\infty} w^2 g(w)\, \mathrm{d}w \\
&\quad \times \int_{\theta=0}^{\pi} \cos^2(\theta)\sin(\theta)\,\mathrm{d}\theta \int_{\varphi=0}^{2\pi} \mathrm{d}\varphi \\
&= \frac{2\pi\mathbf{D}^2}{3V} \int_{\Re^3} g(|\mathbf{w}|)\, \mathrm{d}^3\mathbf{w} \\
&= \frac{2\pi\mathbf{D}^2}{3V} \int_S \mathrm{d}^3\rho \\
&\quad \times \int_{\Re^3} \exp(-2\pi i\mathbf{w}\cdot\rho)\, \mathrm{d}^3\mathbf{w}, \qquad (3.5.2.31)
\end{aligned}$$

where we have switched the order of integration in the last step. Finally, since $\int_{\Re^3} \exp(-2\pi i\mathbf{w}\cdot\rho)\,\mathrm{d}^3\mathbf{w} = \delta_3(\rho)$, the three-dimensional Dirac delta function, this last double integral equals one, completing the derivation of the Ewald correction term in the case of a large but finite crystal having a macroscopically spherical shape.

3.5.2.4. The surface term

Above we have derived the form of the electrostatic energy in the case of a macroscopically spherical crystal in terms of the Ewald sum plus a correction term. In this section, following the approach of van Eijck & Kroon (1997), we develop a surface-integral-based expression for the shape-dependent correction term for more general macroscopic shapes.

Given a closed bounded region P that contains the unit sphere S in its interior and a $K > 0$, and recalling equations (3.5.2.4) and (3.5.2.27), we can write the electrostatic energy of the central unit cell in the finite crystal as

$$\begin{aligned}
E_{P,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= E(U,U) + \tfrac{1}{2} \sum_{\mathbf{n}\in\Omega(P,K),\mathbf{n}\neq\mathbf{0}} E(U, U_{\mathbf{n}}) \\
&= E(U,U) + \tfrac{1}{2} \sum_{\mathbf{n}\in\Omega(S,K),\mathbf{n}\neq\mathbf{0}} E(U, U_{\mathbf{n}}) \\
&\quad + \tfrac{1}{2} \sum_{\mathbf{n}\in\Omega(P,K)\setminus\Omega(S,K)} E(U, U_{\mathbf{n}}) \\
&= E_{S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) + [E_{P,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) - E_{S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})].
\end{aligned}$$

The asymptotic form of $E_{S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})$ has been developed above. Recalling equation (3.5.2.5), for large K and neutral unit cells, we can approximate $\delta E_{P\setminus S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) = [E_{P,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) - E_{S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})]$ by

$$\begin{aligned}
\delta E_{P\setminus S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2} \sum_{(\mathbf{n}/K)\in P\setminus S} E(U, U_{\mathbf{n}}) \\
&\simeq \tfrac{1}{2} \sum_{(\mathbf{n}/K)\in P\setminus S} \frac{\mathbf{D}^2}{|\mathbf{n}|^3} - 3\frac{(\mathbf{D}\cdot\mathbf{n})^2}{|\mathbf{n}|^5} \\
&\simeq \frac{1}{2V} \int_{\mathbf{r}\in P\setminus S} \frac{\mathbf{D}^2}{|\mathbf{r}|^3} - 3\frac{(\mathbf{D}\cdot\mathbf{r})^2}{|\mathbf{r}|^5}\, \mathrm{d}^3\mathbf{r} \\
&\simeq \frac{1}{2V} \oint_{\mathbf{r}\in\partial P\setminus\partial S} \frac{\mathbf{D}\cdot\mathbf{r}}{|\mathbf{r}|^3}\mathbf{D}\cdot\,\mathrm{d}\sigma,
\end{aligned}$$

where we have used Gauss' divergence theorem (Arfken & Weber, 2000) together with

$$\nabla\cdot\left(\frac{\mathbf{D}\cdot\mathbf{r}}{|\mathbf{r}|^3}\mathbf{D}\right) = \frac{\mathbf{D}\cdot\mathbf{D}}{|\mathbf{r}|^3} - 3\frac{(\mathbf{D}\cdot\mathbf{r})^2}{|\mathbf{r}|^5}.$$

Since, considering the outward unit normal on S,

$$\begin{aligned}
\frac{1}{2V} \oint_{\mathbf{r}\in\partial S} \frac{\mathbf{D}\cdot\mathbf{r}}{|\mathbf{r}|^3}\mathbf{D}\cdot\,\mathrm{d}\sigma &= \frac{\mathbf{D}^2}{2V} \int\int \cos^2(\theta)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi \\
&= \frac{2\pi\mathbf{D}^2}{3V}
\end{aligned}$$

and since from the previous results

$$\lim_{K\to\infty} E_{S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) = E_{\mathrm{Ewald}}(\underline{\mathbf{r}}^{\mathbf{N}}) + \frac{2\pi\mathbf{D}^2}{3V},$$

we can now generalize the lattice summation to asymptotic shapes P:

$$\lim_{K\to\infty} E_{S,K}(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) = E_{\mathrm{Ewald}}(\underline{\mathbf{r}}^{\mathbf{N}}) + \frac{1}{2V} \oint_{\mathbf{r}\in\partial P} \frac{\mathbf{D}\cdot\mathbf{r}}{|\mathbf{r}|^3}\mathbf{D}\cdot\mathrm{d}\sigma.$$

The latter surface integral is referred to as the 'surface term' and clearly must equal the quantity $J(\mathbf{D}, P)$ discussed above. It was derived previously by Deem *et al.* (1990) by different arguments. Van Eijck and Kroon calculate it explicitly for some specific crystal shapes. They consider, however, that in actual crystal growth, surface charge distributions should develop to cancel the surface-integral contribution (see also Smith, 1981). In contrast, Scheraga and co-workers argue for the importance of considering this term, which they refer to as the 'Lorentz–Ewald correction', in crystal-structure prediction (Pillardy *et al.*, 2000), although in this they are disputed by van Eijck & Kroon (2000) (see also the response by Wedemeyer *et al.*, 2000).

The results in this section are all derived assuming the crystal is surrounded by vacuum, or at least by a medium with no dielectric screening capability. In the following section we study the effect of a surrounding continuum dielectric medium, assuming a spherical crystal (to make the calculations tractable).

3.5.2.5. The polarization response

In this section, following the approach in Smith (1981, 1994), we derive the polarization correction to the electrostatic energy of U due to immersing the large but finite crystal in a uniform dielectric with dielectric constant $\epsilon > 1$, assuming the macroscopic shape of the crystal is spherical. We first derive the dielectric reaction potential at a point $\mathbf{r} \in U$ due to the charges

$q_1, \ldots, q_N \in U$ and their periodic images in $U_{\mathbf{n}}$, $\mathbf{n} \in \Omega(P, K)$, assuming the unit cell is neutral, *i.e.* $q_1 + \ldots + q_N = 0$, that $P = S$, the unit sphere, and that the sphere KS is immersed in the dielectric.

To begin with, given a point $\mathbf{r} \in \Re^3$ and a point $\mathbf{r}_0 \in KS$, with spherical polar coordinates (r, θ, φ) and $(r_0, \theta_0, \varphi_0)$, respectively, the electrostatic potential at $\mathbf{r}$ due to a unit charge at $\mathbf{r}_0$ is given by the solution to Laplace's equation. Let β be the angle between $\mathbf{r}$ and $\mathbf{r}_0$, that is $\beta = \cos^{-1}(\mathbf{r} \cdot \mathbf{r}_0/|\mathbf{r}||\mathbf{r}_0|)$. Owing to the axial symmetry, the general solution $\psi(\mathbf{r}; \mathbf{r}_0)$ to Laplace's equation in this case can be written

$$\psi(\mathbf{r}; \mathbf{r}_0) = \sum_{l=0}^{\infty} [a_l r^l + b_l r^{-(l+1)}] P_l[\cos(\beta)],$$

where $P_l[\cos(\beta)]$ are the Legendre polynomials (Böttcher, 1973). The contributions a_l can be thought of as due to charges further from the origin than $\mathbf{r}$, while b_l are due to charges closer to the origin. A particular solution to Laplace's equation is given by $1/|\mathbf{r} - \mathbf{r}_0|$. Thus, we can write the potential due to the unit charge at $\mathbf{r}_0$ in the presence of the dielectric continuum surrounding KS as

$$V_1(\mathbf{r}) = \frac{1}{|\mathbf{r} - \mathbf{r}_0|} + \sum_{l=0}^{\infty} \left[A_l r^l + \frac{B_l}{r^{(l+1)}} \right] P_l[\cos(\beta)] \quad \text{for } r \le K,$$

$$V_2(\mathbf{r}) = \sum_{l=0}^{\infty} \left[C_l r^l + \frac{D_l}{r^{(l+1)}} \right] P_l[\cos(\beta)] \quad \text{for } r \ge K.$$

We now apply boundary conditions, noting that P_l are an orthogonal basis of functions, so that boundary conditions can be applied term by term. Then we have

(1) V_1 must be finite as $r \to 0$, implying $B_l = 0$ for all l;

(2) V_2 must go to zero as $r \to \infty$, implying $C_l = 0$ for all $l > 0$.

We expand $1/|\mathbf{r} - \mathbf{r}_0|$ (for $r > r_0$) in Legendre polynomials to allow comparison of V_1 with V_2 as $r \to K$ from below:

$$\frac{1}{|\mathbf{r} - \mathbf{r}_0|} = \sum_{l=0}^{\infty} \frac{r_0^l}{r^{(l+1)}} P_l[\cos(\beta)].$$

Using this we get:

(1) $V_2 = V_1$ for $r = K$, implying $D_l = r_0^l + A_l K^{(2l+1)}$;

(2) $\epsilon(\partial V_2/\partial r) = (\partial V_1/\partial r)$ for $r = K$, implying $A_l = -\{(\epsilon - 1)(l + 1)/[l + \epsilon(l + 1)]\}[r_0^l/K^{(2l+1)}]$.

Finally, since we need the solution for the potential due to each of the charges in $U_{\mathbf{n}}$, we invoke the spherical harmonic addition theorem (Arfken & Weber, 2000) applying it to the points $\mathbf{r}$, $\mathbf{r}_0$:

$$\frac{2l+1}{4\pi} P_l[\cos(\beta)] = \sum_{m=-l}^{+l} (-1)^m Y_l^m(\theta, \varphi) Y_l^{-m}(\theta_0, \varphi_0),$$

where (Arfken & Weber, 2000)

$$Y_l^m(\theta, \varphi) = (-1)^m \left[\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!} \right]^{1/2} P_l^m[\cos(\theta)] \exp(im\varphi)$$

and the associated Legendre polynomials $P_l^m(x)$, $-l \le m \le l$ are defined by Rodrigues' formula,

$$P_l^m[\cos(\theta)] = (-1)^l \frac{1}{2^l l!} \sin(\theta)^m \frac{d^{l+m}}{d[\cos(\theta)]^{l+m}} [1 - \cos^2(\theta)]^l, \tag{3.5.2.32}$$

and

$$P_l^m[\cos(\theta)] = (-1)^m \frac{(l+m)!}{(l-m)!} P_l^{-m}[\cos(\theta)]. \tag{3.5.2.33}$$

In particular, the first-order spherical harmonics are given by

$$Y_1^1(\theta, \varphi) = -\left(\frac{3}{8\pi} \right)^{1/2} \sin(\theta) \exp(i\varphi),$$

$$Y_1^0(\theta, \varphi) = \left(\frac{3}{4\pi} \right)^{1/2} \cos(\theta),$$

$$Y_1^{-1}(\theta, \varphi) = \left(\frac{3}{8\pi} \right)^{1/2} \sin(\theta) \exp(-i\varphi). \tag{3.5.2.34}$$

The spherical harmonics satisfy an orthogonality relation:

$$\int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} [Y_l^m(\theta, \varphi)]^* Y_{l'}^{m'} \sin(\theta) \, \mathrm{d}\theta \, \mathrm{d}\varphi = \delta_{l,l'} \delta_{m,m'}.$$

Then, for $\mathbf{r} \in U$ we express the polarization potential due to the dielectric response to a unit charge at $\mathbf{r}_0 \in KS$ as

$$\psi_{\mathrm{Pol}}(\mathbf{r}; \mathbf{r}_0) = -\sum_{l=0}^{\infty} \sum_{m=-l}^{+l} C(l, m, \epsilon) \frac{1}{K^{2l+1}} g_{l,m}(\mathbf{r}) g_{l,-m}(\mathbf{r}_0),$$

where

$$C(l, m, \epsilon) = (-1)^m \frac{4\pi}{2l+1} \frac{(\epsilon - 1)(l+1)}{l + \epsilon(l+1)}$$

and

$$g_{l,m}(\mathbf{r}) = r^l Y_l^m(\theta, \varphi).$$

We now examine the total polarization potential

$$\begin{aligned}
\psi_{\mathrm{Pol},K,S}(\mathbf{r}) &= \sum_{\mathbf{n} \in \Omega(S,K)} \sum_{j=1}^{N} q_j \psi_{\mathrm{Pol}}(\mathbf{r}; \mathbf{r}_j + \mathbf{n}) \\
&= -\frac{1}{V} \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} C(l, m, \epsilon) \frac{1}{K^{l-1}} g_{l,m}(\mathbf{r}) \cdot K \frac{V}{K^3} \\
&\quad \times \sum_{\mathbf{n} \in \Omega(S,K)} \sum_{j=1}^{N} q_j \left[g_{l,-m}\left(\frac{\mathbf{r}_j + \mathbf{n}}{K} \right) - g_{l,-m}\left(\frac{\mathbf{n}}{K} \right) \right] \\
&= -\frac{1}{V} \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} C(l, m, \epsilon) \frac{1}{K^{l-1}} g_{l,m}(\mathbf{r}) \\
&\quad \times \left[\mathbf{D} \cdot \int_S \nabla g_{l,-m}(\rho) \, \mathrm{d}^3\rho \right] + \varepsilon(K),
\end{aligned} \tag{3.5.2.35}$$

where $\mathbf{D}$ is the unit-cell dipole and we have used unit-cell neutrality in the next to last equation.

The gradient ∇ with respect to the usual Cartesian directions, expressed in spherical polar coordinates (Arfken & Weber, 2000) can be re-expressed as

$$\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}=\sin(\theta)\exp(i\varphi)\frac{\partial}{\partial r}+\frac{1}{r}\cos(\theta)\exp(i\varphi)\frac{\partial}{\partial\theta}+\frac{i\exp(i\varphi)}{r\sin(\theta)}\frac{\partial}{\partial\varphi}$$
$$\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}=\sin(\theta)\exp(-i\varphi)\frac{\partial}{\partial r}+\frac{1}{r}\cos(\theta)\exp(-i\varphi)\frac{\partial}{\partial\theta}-\frac{i\exp(-i\varphi)}{r\sin(\theta)}\frac{\partial}{\partial\varphi}$$
$$\frac{\partial}{\partial z}=\cos(\theta)\frac{\partial}{\partial r}-\frac{1}{r}\sin(\theta)\frac{\partial}{\partial\theta}.$$

Using spherical polar coordinates to carry out the integrations over the unit sphere in equation (3.5.2.35), we see after the first partial integration over φ that

$$\int_S\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right)g_{l,-m}(\rho)\,\mathrm{d}^3\rho=0\quad\text{if } m\neq 1,$$
$$\int_S\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)g_{l,-m}(\rho)\,\mathrm{d}^3\rho=0\quad\text{if } m\neq -1\ \text{ and}$$
$$\int_S\frac{\partial}{\partial z}g_{l,-m}(\rho)\,\mathrm{d}^3\rho=0\quad\text{if } m\neq 0.$$

We recognize the factors multiplying $(\partial/\partial r)$ above as being proportional to Y_1^1, Y_1^0 and Y_1^{-1}, respectively, so that, using the above orthogonality of spherical harmonics, these contribute nothing to the integrals unless $l=1$.

Next, examining I given by

$$I=\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\left[\cos(\theta)\exp(i\varphi)\frac{\partial}{\partial\theta}+\frac{i\exp(i\varphi)}{\sin(\theta)}\frac{\partial}{\partial\varphi}\right]\times g_{l,-1}(r,\theta,\varphi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi,$$

integrating over φ and then applying integration by parts over θ to the first term, expressing $g_{l,-1}(r,\theta,\varphi)$ using the Rodrigues' formulas (3.5.2.32) and (3.5.2.33), and finally substituting $w=\cos(\theta)$ we see that I is proportional to the integral J given by

$$J=\int_{w=-1}^{1}(1-w^2)\frac{d^{l+1}}{dw^{l+1}}(1-w^2)^l\,\mathrm{d}w,$$

which in turn is zero unless $l=1$. Putting the above results together we see that

$$\int_S\left(\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right)g_{l,-m}(\rho)\,\mathrm{d}^3\rho=0\quad\text{unless}\quad l=1,\quad m=1.$$

Similarly

$$\int_S\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)g_{l,-m}(\rho)\,\mathrm{d}^3\rho=0\quad\text{unless}\quad l=1,\quad m=-1.$$

Finally,

$$I=\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\sin(\theta)\frac{\partial}{\partial\theta}g_{l,0}(r,\theta,\varphi)\sin(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi$$

is proportional to

$$J=\int_{w=-1}^{1}w\frac{\mathrm{d}^l}{\mathrm{d}w^l}(1-w^2)^l\,\mathrm{d}w$$

and J is zero unless $l=1$. Thus

$$\int_S\frac{\partial}{\partial z}g_{l,-m}(\rho)\,\mathrm{d}^3\rho=0\quad\text{unless}\quad l=1,\quad m=0.$$

Expressing $g_{1,m}$ in terms of the explicit forms for Y_1^m given in equations (3.5.2.34), we get

$$\int_S\left(\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right)g_{1,1}(\rho)\,\mathrm{d}^3\rho=-(8\pi/3)^{1/2}$$

and thus

$$\int_S\nabla g_{1,1}(\rho)\,\mathrm{d}^3\rho=-(8\pi/3)^{1/2}(1/2,i/2,0).$$

Similarly

$$\int_S\nabla g_{1,0}(\rho)\,\mathrm{d}^3\rho=(4\pi/3)^{1/2}(0,0,1)$$

and finally

$$\int_S\nabla g_{1,-1}(\rho)\,\mathrm{d}^3\rho=(8\pi/3)^{1/2}(1/2,-i/2,0).$$

Substituting these into equation (3.5.2.35) we get

$$\lim_{K\to\infty}\psi_{\mathrm{Pol},K,S}(\mathbf{r})=\frac{-4\pi}{3V}\frac{2\epsilon-2}{2\epsilon+1}(\mathbf{r}\cdot\mathbf{D}).$$

Recalling the shape-dependent term $J(\mathbf{D},S)=2\pi\mathbf{D}^2/3V$ from equation (3.5.2.31) and adding the reaction energy $\sum_{j=1}^{N}q_j\psi_{\mathrm{Pol},K,S}(\mathbf{r}_j)$, we see that the electrostatic energy of the unit cell U inside a large spherical crystal which in turn is immersed in a dielectric continuum with dielectric constant ϵ is given by

$$E_{S,K,\epsilon}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})=E_{\mathrm{Ewald}}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})+\frac{2\pi\mathbf{D}^2}{3V}-\frac{2\pi}{3V}\frac{2\epsilon-2}{2\epsilon+1}\mathbf{D}^2+\varepsilon(K)$$
$$=E_{\mathrm{Ewald}}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})+\frac{1}{2\epsilon+1}\cdot\frac{2\pi}{V}\mathbf{D}^2+\varepsilon(K),$$

where E_{Ewald} is given in equation (3.5.2.15). Note that we recover the usual Ewald sum in the limit as $\epsilon\to\infty$ ('tin-foil' boundary conditions).

The force on a charge q_j in the central cell is obtained by taking the gradient of the energy with respect to $\mathbf{r}_j$ for the finite spherical crystal, and then taking the limit as $K\to\infty$, or by taking the

gradient of the limiting energy above (*i.e.* the limit of the gradient is the gradient of the limit):

$$\mathbf{F}_j = -\nabla_j E_{\rm Ewald}(\underline{\mathbf{r}}^{\mathbf{N}}) - \frac{1}{2\epsilon+1}\cdot\frac{4\pi}{V}q_j\mathbf{D}. \qquad (3.5.2.36)$$

Thus we can find tractable analytic expressions for the energy and force on particles within the central unit cell as long as the asymptotic shape of the crystal immersed in the dielectric medium is spherical. In Section 3.5.2.6 below we show that we can also arrive at a tractable result for the thermodynamic pressure under these circumstances, although surprisingly it does not seem to agree with the mechanical pressure as given by Smith (1994), which may reflect some not yet fully understood subtleties in the implementation of periodic boundary conditions.

3.5.2.6. Calculating the pressure

We first discuss the scalar pressure, following the developments in Smith (1987, 1993). According to the thermodynamic definition, the scalar pressure P is defined by

$$P = -\left(\frac{\partial A}{\partial V}\right)_T,$$

where A is the Helmholtz free energy, V is the volume and T is the temperature. In turn, A is given by $A = -(1/\beta)\ln Z_N(V,T)$ where $\beta = 1/kT$ (k is Boltzmann's constant) and Z_N is the canonical ensemble partition function,

$$Z_N(V,T) = \frac{1}{N!h^{3N}}\int\int \exp[-\beta H_N(\underline{\mathbf{r}}^{\{\mathbf{N}\}},\underline{\mathbf{p}}^{\{\mathbf{N}\}})]\,{\rm d}\underline{\mathbf{r}}^{\{\mathbf{N}\}}\,{\rm d}\underline{\mathbf{p}}^{\{\mathbf{N}\}},$$

where H_N is the system Hamiltonian, h is Planck's constant and $\underline{\mathbf{p}}^{\{\mathbf{N}\}}$ denotes the set of momenta $\{\mathbf{p}_1,\ldots,\mathbf{p}_N\}$ with $\mathbf{p}_i$ conjugate to $\mathbf{r}_i$. Introducing the scaling relations $\mathbf{r}_i = V^{1/3}\mathbf{s}_i$, $\boldsymbol{\pi}_i = V^{2/3}m_i\dot{\mathbf{s}}_i$ and $\mathbf{p}_i = V^{-1/3}\boldsymbol{\pi}_i$, we get

$$\begin{aligned}P &= -\frac{1}{[N!h^{3N}Z_N(V,T)]}\\ &\quad\times\int\int\left[\frac{\partial}{\partial V}H_N(V^{1/3}\underline{\mathbf{s}}^{\{\mathbf{N}\}},V^{-1/3}\underline{\boldsymbol{\pi}}^{\{\mathbf{N}\}})\right]_T\\ &\quad\times\exp[-\beta H_N(V^{1/3}\underline{\mathbf{s}}^{\{\mathbf{N}\}},V^{-1/3}\underline{\boldsymbol{\pi}}^{\{\mathbf{N}\}})]\,{\rm d}\underline{\mathbf{s}}^{\{\mathbf{N}\}}\,{\rm d}\underline{\boldsymbol{\pi}}^{\{\mathbf{N}\}}\\ &= -\left\langle\left[\frac{\partial}{\partial V}H_N(V^{1/3}\underline{\mathbf{s}}^{\{\mathbf{N}\}},V^{-1/3}\underline{\boldsymbol{\pi}}^{\{\mathbf{N}\}})\right]\right\rangle,\end{aligned}$$

where $\langle\ \rangle$ denotes the expectation or ensemble average. For the above system of N point charges $q_1,\ldots,q_N$ having mass $m_1,\ldots,m_N$ at positions $\mathbf{r}_1,\ldots,\mathbf{r}_N$ in the central unit cell U_0 of a large spherical crystal (of radius K) immersed in a dielectric continuum, the Hamiltonian can be written

$$\begin{aligned}H_N(\underline{\mathbf{r}}^{\{\mathbf{N}\}},\underline{\mathbf{p}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2}\sum_{j=1}^{N}\frac{\mathbf{p}_j^2}{m_j} + E_{\rm Ewald}(\underline{\mathbf{r}}^{\{\mathbf{N}\}})\\ &\quad+\frac{1}{2\epsilon+1}\cdot\frac{2\pi}{V}\mathbf{D}^2+\varepsilon(K)\end{aligned}$$

and so

$$\begin{aligned}&H_N(V^{1/3}\underline{\mathbf{s}}^{\{\mathbf{N}\}},V^{-1/3}\underline{\boldsymbol{\pi}}^{\{\mathbf{N}\}})\\ &\quad=\tfrac{1}{2}V^{-2/3}\sum_{j=1}^{N}\frac{\boldsymbol{\pi}_j^2}{m_j}+E_{\rm Ewald}(V^{1/3}\underline{\mathbf{r}}^{\{\mathbf{N}\}})\\ &\qquad+V^{-1/3}\frac{2\pi}{2\epsilon+1}\left(\sum_{j=1}^{N}q_j\mathbf{s}_j\right)^2+\varepsilon(K).\end{aligned}$$

The direct- and reciprocal-lattice vectors are scaled as $\mathbf{n} = V^{1/3}\boldsymbol{\eta}$ and $\mathbf{m} = V^{-1/3}\boldsymbol{\mu}$. The derivative with respect to volume of the Ewald direct sum is straightforward using the chain rule, since if $x = |\mathbf{r}_j+\mathbf{n}| = V^{1/3}|\mathbf{s}_j+\boldsymbol{\eta}|$ then $\partial x/\partial V = x/3V$ and $({\rm d}/{\rm d}x)[{\rm erfc}(\alpha^{1/2}x)/x] = -[(2/\pi^{1/2})\exp(-\alpha x^2)/x + {\rm erfc}(\alpha^{1/2}x)/x^2]$. Since $\mathbf{m}$ and $\mathbf{r}_j$ scale oppositely with V, $\partial S(\mathbf{m})/\partial V = 0$. Continuing with the other terms in the Ewald reciprocal sum, and taking the limit as $K\to\infty$, we get

$$\begin{aligned}P &= \frac{1}{3V}\left\langle\sum_{j=1}^{N}\frac{\mathbf{p}_j^2}{m_j}+\tfrac{1}{2}{\sum_{\mathbf{n}}}'\sum_{i=1}^{N}\sum_{j=1}^{N}q_iq_j\right.\\ &\quad\times\left\{\frac{2}{\pi^{1/2}}\exp\left[-\alpha(\mathbf{r}_j-\mathbf{r}_i+\mathbf{n})^2\right]+\frac{{\rm erfc}(\alpha^{1/2}|\mathbf{r}_j-\mathbf{r}_i+\mathbf{n}|)}{|\mathbf{r}_j-\mathbf{r}_i+\mathbf{n}|}\right\}\\ &\quad+\frac{1}{2\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}\left(1-\frac{2\pi\mathbf{m}^2}{\alpha}\right)S(\mathbf{m})S(-\mathbf{m})\\ &\quad\left.+\frac{1}{2\epsilon+1}\cdot\frac{2\pi}{V}\mathbf{D}^2\right\rangle.\end{aligned}$$

The scalar pressure has been generalized to the pressure tensor (Brown & Neyertz, 1995) using a more general scaling technique based on the expression for $\mathbf{r}$ in terms of unit-cell basis vectors and fractional coordinates, treating the standard Ewald sum without the surface correction or dielectric response term. To accommodate the latter, the energy of the central unit cell within a large ellipsoidal crystal immersed in a dielectric continuum must be studied. In Redlack & Grindley (1972, 1975) expressions are provided for the surface correction in ellipsoidal crystals in a vacuum. If this result were combined with the polarization response term (we are not aware of any papers giving this term for an ellipsoidal crystal immersed in a dielectric continuum) it is reasonable to assume that the standard Ewald sum would apply in this case as well, in the limits $K\to\infty$ and $\epsilon\to\infty$. The trace of the pressure tensor defined by Nosé & Klein (1983) agrees with the above scalar pressure in the limit $\epsilon\to\infty$.

Interestingly, Smith (1994) derives a scalar pressure that differs from the above. In particular he finds a surface term that remains even in the limit of 'tin-foil' boundary conditions, *i.e.* when $\epsilon\to\infty$. He gives a mechanical definition of the pressure tensor in terms of momentum flux across surfaces within and surrounding the crystal. Periodic boundary conditions are enforced using velocity constraints *via* the techniques of DeLeeuw *et al.* (1990). He derives forces on particles within the central unit cell (and on their images in other cells interior to the crystal) that agree in the large K limit with our result above in equation (3.5.2.36). However, combining the forces with particle positions in the scalar virial expression, after summing over particles and their images, and proceeding to the limit as $K\to\infty$ leads to a vacuum surface correction involving the reciprocal sum term at $\mathbf{m} = \mathbf{0}$ that, unlike our thermodynamic pressure expression above, is not cancelled by the dielectric response potential in the limit $\epsilon\to\infty$. To reiterate, his mechanical virial tensor, restricted to the standard Ewald sum part of the final expression, agrees completely with the results of Nosé & Klein (1983). Furthermore, his scalar virial, given by the trace of his virial tensor, has an ϵ-dependent part due to the dielectric response

that agrees completely with our results above. However, the trace of the vacuum surface correction tensor within his mechanical pressure tensor does not agree with the equivalent term in our thermodynamic pressure derived above, and thus is not cancelled in the infinite dielectric limit.

3.5.3. Generalization to Gaussian- and Hermite-based continuous charge distributions

In this section we develop the generalization of Ewald sums to interacting molecular densities modelled by linear combinations of spherical or Hermite Gaussians. In Coppens (1997) the electrostatic potential and energy for crystal densities modelled by linear combinations of Slater-like density elements is discussed. As with quantum-chemical programs, there are trade-offs in the choice of density elements. The use of Slater-like densities provides accuracy with the use of fewer elements, but the Gaussian and Hermite Gaussian elements are far more convenient analytically. In particular the Ewald summation, as well as the fast-Fourier-transform-based methods discussed in Section 3.5.4, generalize straightforwardly to these elements. We note ahead of time that the case of lattice sums of point multipoles are obtained by specializing the results for Hermite Gaussians, just as the results for spherical Gaussians specialize to the above lattice summation results for point charges.

A somewhat different approach to Ewald summation over Hermite Gaussian elements, focused on application to periodic density-functional calculations, is given in Trickey *et al.* (2004). The continuous fast multipole method (Challacombe *et al.*, 1997) provides yet a different but highly efficient approach to lattice summation of continuous charge densities.

3.5.3.1. Lattice sums of interacting spherical Gaussians

We begin by discussing the electrostatic interaction energy between a normalized spherical Gaussian charge distribution ρ_1, centred at a point $\mathbf{R}_1 \in U$ and having exponent θ_1, and a second normalized spherical Gaussian charge distribution ρ_2, centred at $\mathbf{R}_2 \in U$ and having exponent θ_2, together with the latter's periodic images, centred at $\mathbf{R}+\mathbf{n}$, $\mathbf{n} \in \Omega(S,K)$. Their interaction energy can be written

$$E_{12} = \sum_{\mathbf{n}\in\Omega(S,K)} \int_{\Re^3}\int_{\Re^3} \rho_1(\mathbf{r}_1-\mathbf{R}_1)\rho_2(\mathbf{r}_2-\mathbf{R}_2-\mathbf{n}) \times 1/|\mathbf{r}_2-\mathbf{r}_1|\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2, \tag{3.5.3.1}$$

where $\rho_j(\mathbf{r}) = (\theta_j/\pi)^{3/2}\exp(-\theta_j\mathbf{r}^2)$, $j = 1, 2$. We now recall equation (3.5.2.16), applying its two right-hand components separately, to get $E_{12} = E_{12}^{\mathrm{d}} + E_{12}^{\mathrm{r}}$, where

$$E_{12}^{\mathrm{d}} = \sum_{\mathbf{n}\in\Omega(S,K)} \int_{\Re^3}\int_{\Re^3} \rho_1(\mathbf{r}_1-\mathbf{R}_1)\rho_2(\mathbf{r}_2-\mathbf{R}_2-\mathbf{n}) \times \frac{\mathrm{erfc}(\alpha^{1/2}|\mathbf{r}_2-\mathbf{r}_1|)}{|\mathbf{r}_2-\mathbf{r}_1|}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2$$

and

$$E_{12}^{\mathrm{r}} = \frac{1}{\pi}\sum_{\mathbf{n}\in\Omega(S,K)} \int_{\Re^3}\int_{\Re^3} \rho_1(\mathbf{r}_1-\mathbf{R}_1)\rho_2(\mathbf{r}_2-\mathbf{R}_2-\mathbf{n}) \times \int_{\Re^3} \frac{\exp(-\pi^2\mathbf{u}^2/\alpha)}{\mathbf{u}^2}\exp[-2\pi i\mathbf{u}\cdot(\mathbf{r}_2-\mathbf{r}_1)]\,\mathrm{d}^3\mathbf{u}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2.$$

The direct pair energy E_{12}^{d} involves the damped Coulomb interaction, which has been discussed by Gill & Adamson (1996). We can calculate E_{12}^{d} using the following standard results from quantum chemistry, which are derived in Helgaker *et al.* (2000): Let $\rho_p(\mathbf{r})$, $\rho_q(\mathbf{r})$ be two normalized spherical Gaussian charge distributions having exponents p and q and centred at points $\mathbf{P}$ and $\mathbf{Q}$, *i.e.* $\rho_p(\mathbf{r}) = (p/\pi)^{3/2}\exp[-p(\mathbf{r}-\mathbf{P})^2]$ and $\rho_q(\mathbf{r}) = (q/\pi)^{3/2}\exp[-q(\mathbf{r}-\mathbf{Q})^2]$. Then the electrostatic potential $V_p(\mathbf{R})$ at a point $\mathbf{R}$ due to the charge distribution $\rho_p(\mathbf{r})$ is given by

$$V_p(\mathbf{R}) = \left(\frac{4p}{\pi}\right)^{1/2} F_0[p(\mathbf{P}-\mathbf{R})^2], \tag{3.5.3.2}$$

where F_0 is the zeroth-order Boy's function, with $F_0(x) = \int_0^1 \exp(-xt^2)\,\mathrm{d}t$ for $x > 0$. The Coulomb interaction energy E_{pq} between ρ_p and ρ_q is given by

$$\begin{aligned} E_{pq} &= \int_{\Re^3}\int_{\Re^3} \frac{\rho_p(\mathbf{r}_1-\mathbf{P})\rho_q(\mathbf{r}_2-\mathbf{Q})}{|\mathbf{r}_1-\mathbf{r}_2|}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2 \\ &= \int_{\Re^3} V_p(\mathbf{r}_2)\rho_q(\mathbf{r}_2-\mathbf{Q})\,\mathrm{d}^3\mathbf{r}_2 \\ &= \left(\frac{4\mu_{pq}}{\pi}\right)^{1/2} F_0[\mu_{pq}(\mathbf{P}-\mathbf{Q})^2], \end{aligned} \tag{3.5.3.3}$$

where $1/\mu_{pq} = 1/p + 1/q$. The Boy's function can be related to the error function:

$$\frac{2}{\pi^{1/2}}F_0(x^2) = \frac{\mathrm{erf}(|x|)}{|x|}. \tag{3.5.3.4}$$

Substituting equation (3.5.3.4) into (3.5.3.2), we see that

$$V_p(\mathbf{R}) = \frac{\mathrm{erf}(p^{1/2}|\mathbf{R}-\mathbf{P}|)}{|\mathbf{R}-\mathbf{P}|}, \tag{3.5.3.5}$$

a result which we had derived above by another method [see equation (3.5.2.7)]. From equations (3.5.3.3) and (3.5.3.5), substituting α for p and ρ_p for ρ_q, we have for any point $\mathbf{r} \in \Re^3$

$$\int_{\Re^3} \frac{\mathrm{erf}(\alpha^{1/2}|\mathbf{r}-\mathbf{r}_1|)}{|\mathbf{r}-\mathbf{r}_1|}\rho_p(\mathbf{r}_1-\mathbf{P})\,\mathrm{d}^3\mathbf{r}_1 = \frac{\mathrm{erf}(\mu_{p\alpha}^{1/2}\,|\mathbf{r}-\mathbf{P}|)}{|\mathbf{r}-\mathbf{P}|},$$

where $1/\mu_{p\alpha} = 1/p + 1/\alpha$. Iterating this result with $\mu_{p\alpha}$ replacing α and ρ_q replacing ρ_p we get the damped interaction energy between ρ_p and ρ_q,

$$
\begin{aligned}
E_{pq}^{\mathrm{d}} &= \int_{\Re^3}\int_{\Re^3} \rho_p(\mathbf{r}_1-\mathbf{P})\rho_q(\mathbf{r}_2-\mathbf{Q}) \\
&\quad\times \frac{\operatorname{erfc}(\alpha^{1/2}|\mathbf{r}_2-\mathbf{r}_1|)}{|\mathbf{r}_2-\mathbf{r}_1|}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2 \\
&= \int_{\Re^3}\int_{\Re^3} \frac{\rho_p(\mathbf{r}_1-\mathbf{P})\rho_q(\mathbf{r}_2-\mathbf{Q})}{|\mathbf{r}_1-\mathbf{r}_2|}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2 \\
&\quad - \int_{\Re^3}\int_{\Re^3} \rho_p(\mathbf{r}_1-\mathbf{P})\rho_q(\mathbf{r}_2-\mathbf{Q}) \\
&\quad\times \frac{\operatorname{erf}(\alpha^{1/2}|\mathbf{r}_2-\mathbf{r}_1|)}{|\mathbf{r}_2-\mathbf{r}_1|}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2 \\
&= \frac{\operatorname{erf}(\mu_{pq}^{1/2}|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}-\mathbf{Q}|} - \frac{\operatorname{erf}(\mu_{pq\alpha}^{1/2}|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}-\mathbf{Q}|} \\
&= \frac{\operatorname{erfc}(\mu_{pq\alpha}^{1/2}|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}-\mathbf{Q}|} - \frac{\operatorname{erfc}(\mu_{pq}^{1/2}|\mathbf{P}-\mathbf{Q}|)}{|\mathbf{P}-\mathbf{Q}|},
\end{aligned}
\tag{3.5.3.6}
$$

where $1/\mu_{pq} = 1/p + 1/q$ and $1/\mu_{pq\alpha} = 1/p + 1/q + 1/\alpha$. Notice that in the limit $\alpha \to \infty$, $E_{pq,\alpha} \to 0$ as expected, since $\operatorname{erfc}(x) \to 0$ as $x \to \infty$, and that in the limit $p \to \infty$, $q \to \infty$, $E_{pq,\alpha} \to \operatorname{erfc}(\alpha^{1/2}|\mathbf{P}-\mathbf{Q}|)/|\mathbf{P}-\mathbf{Q}|$, the damped interaction between two unit point charges at $\mathbf{P}$ and $\mathbf{Q}$, as in the standard Ewald direct sum.

To calculate the reciprocal pair energy E_{12}^{r} we first apply the first right-hand component of equation (3.5.2.16) together with the Fourier transforms of the normalized Gaussians ρ_p and ρ_q from equation (3.5.2.18) to obtain

$$
\begin{aligned}
E_{pq}^{\mathrm{r}} &= \frac{1}{\pi}\int_{\Re^3}\int_{\Re^3}\int_{\Re^3} \rho_p(\mathbf{r}_1-\mathbf{P})\rho_q(\mathbf{r}_2-\mathbf{Q})\frac{\exp(-\pi^2\mathbf{u}^2/\alpha)}{\mathbf{u}^2} \\
&\quad\times \exp[-2\pi i\mathbf{u}\cdot(\mathbf{r}_1-\mathbf{r}_2)]\,\mathrm{d}^3\mathbf{u}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2 \\
&= \frac{1}{\pi}\int_{\Re^3}\int_{\Re^3}\int_{\Re^3} \frac{1}{\mathbf{u}^2}\exp[-\pi^2(\mathbf{u}^2/\alpha + \mathbf{v}^2/p + \mathbf{w}^2/q)] \\
&\quad\times \exp(2\pi i\mathbf{v}\cdot\mathbf{P})\exp(2\pi i\mathbf{w}\cdot\mathbf{Q})\,\mathrm{d}^3\mathbf{u}\,\mathrm{d}^3\mathbf{v}\,\mathrm{d}^3\mathbf{w} \\
&\quad\times \int_{\Re^3}\exp[-2\pi i(\mathbf{u}+\mathbf{v})\cdot\mathbf{r}_1]\,\mathrm{d}^3\mathbf{r}_1 \\
&\quad\times \int_{\Re^3}\exp[2\pi i(\mathbf{u}-\mathbf{w})\cdot\mathbf{r}_2]\,\mathrm{d}^3\mathbf{r}_2 \\
&= \frac{1}{\pi}\int_{\Re^3}\frac{\exp(-\pi^2\mathbf{u}^2/\mu_{pq\alpha})}{\mathbf{u}^2}\exp[-2\pi i\mathbf{u}\cdot(\mathbf{P}-\mathbf{Q})]\,\mathrm{d}^3\mathbf{u},
\end{aligned}
\tag{3.5.3.7}
$$

where we have used the fact that $\int \exp[-2\pi i(\mathbf{u}+\mathbf{v})\cdot\mathbf{r}_1]\,\mathrm{d}^3\mathbf{r}_1 = \delta^3(\mathbf{u}+\mathbf{v})$, with δ^3 the three-dimensional delta function, and similarly for the integral over $\mathbf{r}_2$. Note that in the limit $p \to \infty$, $q \to \infty$, E_{pq}^{r} reduces to the equivalent expression for unit point charges at $\mathbf{P}$ and $\mathbf{Q}$, and in the limit $\alpha \to \infty$, E_{pq}^{r} reduces to the Fourier-space expression for the direct Coulomb interaction between the Gaussians ρ_p and ρ_q.

Recalling the developments leading to equation (3.5.2.28), and using equations (3.5.3.6) and (3.5.3.7), we can write the interaction energy E_{12} from equation (3.5.3.1) as

$$
\begin{aligned}
E_{12} &= \sum_{\mathbf{n}} \frac{\operatorname{erfc}(\mu_{12\alpha}^{1/2}|\mathbf{R}_{12}-\mathbf{n}|) - \operatorname{erfc}(\mu_{12}^{1/2}|\mathbf{R}_{12}-\mathbf{n}|)}{|\mathbf{R}_{12}-\mathbf{n}|} \\
&\quad + \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}} \frac{\exp(-\pi^2\mathbf{m}^2/\mu_{12\alpha})}{\mathbf{m}^2}\exp[-2\pi i\mathbf{m}\cdot(\mathbf{R}_{12})] \\
&\quad - \frac{\pi}{\mu_{12\alpha}V} + \frac{1}{\pi}H_{S,K}(\mathbf{R}_{12}) + \varepsilon(K),
\end{aligned}
\tag{3.5.3.8}
$$

where $\mathbf{R}_{12} = \mathbf{R}_1 - \mathbf{R}_2$, $1/\mu_{12} = 1/\theta_1 + 1/\theta_2$ and $1/\mu_{12\alpha} = 1/\theta_1 + 1/\theta_2 + 1/\alpha$.

Note that although the energy of a finite-exponent Gaussian charge distribution interacting with itself is finite, when considering a generalization of lattice sums involving point charges it is necessary to remove the direct Coulomb interaction of the Gaussian with itself. From equations (3.5.3.3) and (3.5.3.6), if $p = q$, then as $\mathbf{Q} \to \mathbf{P}$, $E_{pq}^{\mathrm{d}} - E_{pq} \to -2(\mu_{pq\alpha}/\pi)^{1/2}$. We can now generalize equation (3.5.2.29). Let $\rho_1, \ldots, \rho_N$ be normalized spherical Gaussian charge distributions with Gaussian exponents $\theta_1, \ldots, \theta_N$, centred at $\{\mathbf{R}_1, \ldots, \mathbf{R}_N\} = \underline{\mathbf{R}}^{\{\mathbf{N}\}} \in U$ (point charges are included by taking $\theta \to \infty$), and let $q_1, \ldots, q_N$ satisfy $q_1 + \ldots + q_N = 0$ (neutral unit cell). Then the energy of the central unit cell U within a large spherical crystal, due to interactions of the Gaussian charge distributions $q_i\rho_i$ with each other and with their periodic images within the crystal, is given by

$$
\begin{aligned}
E_{S,K}(\underline{\mathbf{R}}^{\{\mathbf{N}\}}) &= \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{i,j=1}^{N} q_iq_j\left\{\frac{\operatorname{erfc}(\mu_{ij\alpha}^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|)}{|\mathbf{R}_{ij}-\mathbf{n}|} - \frac{\operatorname{erfc}(\mu_{ij}^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|)}{|\mathbf{R}_{ij}-\mathbf{n}|}\right\} \\
&\quad + \frac{1}{2\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\sum_{i,j=1}^{N} q_iq_j\frac{\exp(-\pi^2\mathbf{m}^2/\mu_{ij\alpha})}{\mathbf{m}^2} \\
&\quad\times \exp(-2\pi i\mathbf{m}\cdot\mathbf{R}_{ij}) \\
&\quad - \frac{\pi}{2V}\sum_{i,j=1}^{N}\frac{q_iq_j}{\mu_{ij\alpha}} - \sum_{i=1}^{N} q_i^2\left(\frac{\mu_{ii\alpha}}{\pi}\right)^{1/2} \\
&\quad + \frac{2\pi\mathbf{D}^2}{3V} + \varepsilon(K),
\end{aligned}
$$

where $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$, $1/\mu_{ij} = 1/\theta_i + 1/\theta_j$, $1/\mu_{ij\alpha} = 1/\theta_i + 1/\theta_j + 1/\alpha$ and as before $\mathbf{D} = q_1\mathbf{R}_1 + \ldots + q_N\mathbf{R}_N$ is the unit-cell dipole.

Note that from the above derivation, the α in $\mu_{ij\alpha}$ is allowed to be different for each pair ij. One choice that leads to an efficient algorithm (particularly when combined with the fast-Fourier-transform-based methods discussed below) is to separate the Gaussian charge distributions $q_j\rho_j$ into compact and diffuse sets, based on their exponents θ_j, and choose α to be infinite for ij pairs where at least one of the two Gaussians is diffuse. These pair interactions are then evaluated entirely in reciprocal space. Next choose α so that $\mu_{ij\alpha}$ is constant for all compact pairs ij. More specifically, given $\beta > 0$, a Gaussian $q_j\rho_j$ is classified as compact if $\theta_j \geq 2\beta$ and diffuse otherwise. We denote these cases as $j \in C$ and $j \in D$, respectively. Then for $i, j \in C$, choose α so that $1/\mu_{ij\alpha} = 1/\theta_i + 1/\theta_j + 1/\alpha = 1/\beta$. Otherwise, choose α to be infinite. Then the Coulomb energy of the central unit cell $E_{S,K}(\underline{\mathbf{R}}^{\{\mathbf{N}\}})$ can be re-expressed as

$$E_{S,K}(\underline{\mathbf{R}}^{\{\mathbf{N}\}}) = \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{(i,j)\in C\times C} q_i q_j \left\{ \frac{\operatorname{erfc}(\beta^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|)}{|\mathbf{R}_{ij}-\mathbf{n}|} - \frac{\operatorname{erfc}(\mu_{ij}^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|)}{|\mathbf{R}_{ij}-\mathbf{n}|} \right\}$$
$$+ \frac{1}{2\pi V} \sum_{\mathbf{m}\neq\mathbf{0}} \sum_{(i,j)\in C\times C} q_i q_j \, \frac{\exp(-\pi^2\mathbf{m}^2/\beta)}{\mathbf{m}^2} \times \exp(-2\pi i \mathbf{m}\cdot\mathbf{R}_{ij})$$
$$+ \frac{1}{2\pi V} \sum_{\mathbf{m}\neq\mathbf{0}} \sum_{(i,j)\notin C\times C} q_i q_j \, \frac{\exp(-\pi^2\mathbf{m}^2/\mu_{ij})}{\mathbf{m}^2} \times \exp(-2\pi i \mathbf{m}\cdot\mathbf{R}_{ij})$$
$$- \frac{\pi}{2V} \sum_{(i,j)\in C\times C} q_i q_j \left(\frac{1}{\beta} - \frac{1}{\theta_i} - \frac{1}{\theta_j}\right) - \sum_{i\in C} q_i^2 \left(\frac{\beta}{\pi}\right)^{1/2} - \sum_{i\notin C} q_i^2 \left(\frac{\theta_i}{2\pi}\right)^{1/2} + \frac{2\pi\mathbf{D}^2}{3V} + \varepsilon(K). \tag{3.5.3.9}$$

Note that in the limit that $\theta_i \to \infty$ for all i, all of the Gaussians become compact and this formula reduces to equation (3.5.2.30) for a large spherical crystal of point charges. The volume-dependent term $(\pi/2V)\sum_{(i,j)\in C\times C} q_i q_j (1/\beta - 1/\theta_i - 1/\theta_j)$ is somewhat surprising, and would not appear in a more straightforward Ewald derivation such as our first Ewald derivation. It vanishes by unit-cell neutrality for the case of point charges or multipoles, and would also vanish if we chose a single α for all Gaussian pairs.

3.5.3.2. Higher angular momentum: Hermite Gaussians

More complex crystal charge distributions can be realized by considering normalized Hermite Gaussian distributions in place of the normalized spherical Gaussian distributions discussed above. Let ρ be a normalized spherical Gaussian charge distribution with exponent θ, centred at $\mathbf{R} = (\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z)$, *i.e.* $\rho(\mathbf{r}) = (\theta/\pi)^{3/2}\exp[-\theta(\mathbf{r}-\mathbf{R})^2]$. The associated normalized Hermite Gaussians Λ_{tuv}, where t, u, v are non-negative integers, are defined by

$$\Lambda_{tuv}(\mathbf{r},\theta,\mathbf{R}) = \left(\frac{\theta}{\pi}\right)^{3/2} (\partial/\partial\mathbf{R}_x)^t \left(\partial/\partial\mathbf{R}_y\right)^u \left(\partial/\partial\mathbf{R}_z\right)^v \times \exp[-\theta(\mathbf{r}-\mathbf{R})^2].$$

Note that this definition of Hermite Gaussians differs from that found in some textbooks, *e.g.* Helgaker *et al.* (2000), in that we have normalized the Gaussian ρ it is derived from in order to allow exponents to tend to infinity to arrive at ideal point charges or point multipoles. We will refer to these Hermite Gaussians as normalized Hermite Gaussians to emphasize this distinction. The Hermite definition reduces to the spherical Gaussian when t, u and v are all zero, *i.e.* $\Lambda_{000}(\mathbf{r},\theta,\mathbf{R}) = \rho(\mathbf{r})$.

Above we discussed the interaction between spherical Gaussian charge distributions ρ. Here we generalize these to higher angular momentum charge distributions ϱ given by

$$\varrho(\mathbf{r},\theta,\mathbf{R},\mathbf{c}) = \sum_{tuv} \mathbf{c}_{tuv}\Lambda_{tuv}(\mathbf{r},\theta,\mathbf{R}),$$

where $\mathbf{c}$ is a vector of coefficients, a finite number of which are nonzero. For example, if $\mathbf{c}_{000}$ is the only nonzero coefficient, we are reduced to the spherical charge distributions discussed above.

The fact that the partial derivatives within Λ_{tuv} are with respect to the position of the centre of the distribution allows us to easily calculate the Coulomb interaction energy between two normalized Hermite Gaussian distributions. The partial derivatives can be pulled out of the double integral, so that the interaction energy is given by the corresponding partial derivatives of the interaction energy of the two spherical Gaussian charge distributions:

$$E_{12} = \int_{\Re^3}\int_{\Re^3} \frac{\Lambda_{t_1u_1v_1}(\mathbf{r}_1,\theta_1,\mathbf{R}_1)\Lambda_{t_2u_2v_2}(\mathbf{r}_2,\theta_2,\mathbf{R}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2$$
$$= \left(\frac{\partial}{\partial\mathbf{R}_{1x}}\right)^{t_1}\left(\frac{\partial}{\partial\mathbf{R}_{1y}}\right)^{u_1}\left(\frac{\partial}{\partial\mathbf{R}_{1z}}\right)^{v_1}\left(\frac{\partial}{\partial\mathbf{R}_{2x}}\right)^{t_2} \times \left(\frac{\partial}{\partial\mathbf{R}_{2y}}\right)^{u_2}\left(\frac{\partial}{\partial\mathbf{R}_{2z}}\right)^{v_2}\left(\frac{\theta_1\theta_2}{\pi^2}\right)^{3/2}\int_{\Re^3}\int_{\Re^3}\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|} \times \exp[-\theta_1(\mathbf{r}_1-\mathbf{R}_1)^2]\exp[-\theta_2(\mathbf{r}_2-\mathbf{R}_2)^2]\,\mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2.$$

Next we consider the interaction energy between a generalized charge distribution ϱ_1 centred at $\mathbf{R}_1 \in U$ and another generalized charge distribution ϱ_2 centred at $\mathbf{R}_2 \in U$, together with all the latter's periodic images, centred at $\mathbf{R}_2 + \mathbf{n}$, $\mathbf{n} \in \Omega(S,K)$. Examining equation (3.5.3.8), we note that the terms to be differentiated all depend on $\mathbf{R}_{12} = \mathbf{R}_1 - \mathbf{R}_2$. Thus

$$E_{12} = \sum_{\mathbf{n}}\sum_{t_1u_1v_1}\sum_{t_2u_2v_2}(-1)^{(t_2+u_2+v_2)}\mathbf{c}_{1,t_1u_1v_1}\mathbf{c}_{2,t_2u_2v_2} \times \left(\frac{\partial}{\partial\mathbf{R}_{12x}}\right)^{t_1+t_2}\left(\frac{\partial}{\partial\mathbf{R}_{12y}}\right)^{u_1+u_2}\left(\frac{\partial}{\partial\mathbf{R}_{12z}}\right)^{v_1+v_2} \times \left\{\frac{\operatorname{erfc}(\mu_{12\alpha}^{1/2}|\mathbf{R}_{12}-\mathbf{n}|)}{|\mathbf{R}_{12}-\mathbf{n}|} - \frac{\operatorname{erfc}(\mu_{12}^{1/2}|\mathbf{R}_{12}-\mathbf{n}|)}{|\mathbf{R}_{12}-\mathbf{n}|}\right\}$$
$$+ \frac{1}{\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\mu_{12\alpha})}{\mathbf{m}^2}C_1(\mathbf{m})C_2(-\mathbf{m}) - \mathbf{c}_{1,000}\mathbf{c}_{2,000}\frac{\pi}{\mu_{p12\alpha}V} + \frac{1}{\pi}H_{S,K}^{12}(\mathbf{R}_{12}) + \varepsilon(K), \tag{3.5.3.10}$$

where $C_i(\mathbf{m})$, i =1, 2 are given by

$$C_i(\mathbf{m}) = \sum_{t_iu_iv_i}\mathbf{c}_{i,t_iu_iv_i} \times \left(\frac{\partial}{\partial\mathbf{R}_{ix}}\right)^{t_i}\left(\frac{\partial}{\partial\mathbf{R}_{iy}}\right)^{u_i}\left(\frac{\partial}{\partial\mathbf{R}_{iz}}\right)^{v_i}\exp(-2\pi i\mathbf{m}\cdot\mathbf{R}_i)$$

and

$$H_{S,K}^{12}(\mathbf{R}_{12}) = \sum_{t_1u_1v_1}\sum_{t_2u_2v_2}(-1)^{(t_2+u_2+v_2)}\mathbf{c}_{1,t_1u_1v_1}\mathbf{c}_{2,t_2u_2v_2} \times \left(\frac{\partial}{\partial\mathbf{R}_{12x}}\right)^{t_1+t_2}\left(\frac{\partial}{\partial\mathbf{R}_{12y}}\right)^{u_1+u_2}\left(\frac{\partial}{\partial\mathbf{R}_{12z}}\right)^{v_1+v_2} \times H_{S,K}(\mathbf{R}_{12}).$$

Note that by the arguments leading to equation (3.5.2.28), $H_{S,K}(\mathbf{r})$ is quadratic in $\mathbf{r}$, *i.e.* derivatives of higher than second order vanish.

We now can discuss the Coulomb energy of the central unit cell U within a large spherical crystal when the charge density in the

unit cell is described by a basis Θ of normalized Hermite Gaussians and a set of expansion points $\mathbf{R}_1, \ldots, \mathbf{R}_N$ within U that are repeated periodically in the crystal (for example, some of the expansion points could be atomic nuclei). Let $\theta_1, \ldots, \theta_L$ be positive Gaussian exponents (or infinite, in the case of nuclear charge or point charges or ideal multipoles) and let the charge density about the expansion point $\mathbf{R}_i$ be given by

$$\varrho_i(\mathbf{r}, \mathbf{R}_i, \Theta) = \sum_{l=1}^{L} \sum_{tuv} \mathbf{c}_{i,l,tuv} \Lambda_{tuv}(\mathbf{r}, \theta_l, \mathbf{R}_i),$$

where, as above, for each i, l the coefficients $\mathbf{c}_{i,l,tuv}$ are nonzero for a finite number of tuv. As above we can separate the exponents θ_l into compact and diffuse, *i.e.* $l \in C$ or $l \in D$, according to whether $\theta_l \geq 2\beta$ or $\theta < 2\beta$, respectively. Let $q_i = \sum_{l=1}^{L} \mathbf{c}_{i,l,000}$ denote the net charge at the expansion point $\mathbf{R}_i$.

We are interested in the Coulomb energy of charge densities $\varrho_i(\mathbf{r}, \mathbf{R}_i, \Theta)$, $i = 1, \ldots, N$, interacting with each other and their periodic images at expansion points $\mathbf{R}_i + \mathbf{n}, \mathbf{n} \in \Omega(S, K)$ for large K. As above, the direct Coulomb interaction of $\varrho_i(\mathbf{r}, \mathbf{R}_i, \Theta)$ with itself is not permitted. This can be modified for more sophisticated treatments *e.g.* in the context of the Coulomb integrals in a periodic density-functional-theory approach, but clearly a nuclear charge cannot interact with itself. The unit cell is assumed to be neutral, *i.e.* $\sum_{i=1}^{N} q_i = 0$. Then, using equations (3.5.3.9) and (3.5.3.10) we can write the Coulomb energy of $\underline{\varrho}^{\{\mathbf{N}\}} = \{\varrho_1, \ldots, \varrho_N\}$ within the spherical crystal as

$$\begin{aligned}
E_{S,K}(\underline{\varrho}^{\{\mathbf{N}\}}) = {} & \tfrac{1}{2}{\sum_{\mathbf{n}}}' \sum_{i=1}^{N} \sum_{l_i \in C} \sum_{t_i u_i v_i} \mathbf{c}_{i,l_i,t_i u_i v_i} \\
& \times \sum_{j=1}^{N} \sum_{l_j \in C} \sum_{t_j u_j v_j} (-1)^{(t_j+u_j+v_j)} \mathbf{c}_{j,l_j,t_j u_j v_j} \\
& \times \left(\frac{\partial}{\partial \mathbf{R}_{ijx}}\right)^{t_i+t_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijy}}\right)^{u_i+u_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijz}}\right)^{v_i+v_j} \\
& \times \left\{ \frac{\operatorname{erfc}(\beta^{1/2}|\mathbf{R}_{ij} - \mathbf{n}|)}{|\mathbf{R}_{ij} - \mathbf{n}|} - \frac{\operatorname{erfc}(\mu_{l_i l_j}^{1/2}|\mathbf{R}_{ij} - \mathbf{n}|)}{|\mathbf{R}_{ij} - \mathbf{n}|} \right\} \\
& + \frac{1}{2\pi V} \sum_{\mathbf{m} \neq \mathbf{0}} \frac{1}{\mathbf{m}^2} \exp(-\pi^2 \mathbf{m}^2/2\beta) \sum_{l_1 \in C} S_{l_1}(\mathbf{m}) \\
& \times \exp(-\pi^2 \mathbf{m}^2/2\beta) \sum_{l_2 \in C} S_{l_2}(-\mathbf{m}) \\
& + \frac{1}{2\pi V} \sum_{\mathbf{m} \neq \mathbf{0}} \frac{1}{\mathbf{m}^2} \sum_{(l_1,l_2) \notin C \times C} \exp(-\pi^2 \mathbf{m}^2/\theta_{l_1}) \\
& \times \exp(-\pi^2 \mathbf{m}^2/\theta_{l_2}) S_{l_1}(\mathbf{m}) S_{l_2}(-\mathbf{m}) \\
& - \frac{\pi}{2V} \sum_{l_1 \in C} \sum_{l_2 \in C} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{c}_{i,l_1,000} \mathbf{c}_{j,l_2,000} \\
& \times \left(\frac{1}{\beta} - \frac{1}{\theta_{l_1}} - \frac{1}{\theta_{l_2}}\right) \\
& - \sum_{i=1}^{N} E_{\text{self}}(\varrho_i) + \frac{2\pi \mathbf{D}^2}{3V} + \varepsilon(K), \qquad (3.5.3.11)
\end{aligned}$$

where the structure factors $S_l(\mathbf{m})$ are given by

$$\begin{aligned}
S_l(\mathbf{m}) = {} & \sum_{i=1}^{N} \sum_{tuv} \mathbf{c}_{i,l,tuv} \left(\frac{\partial}{\partial \mathbf{R}_{ix}}\right)^{t} \left(\frac{\partial}{\partial \mathbf{R}_{iy}}\right)^{u} \left(\frac{\partial}{\partial \mathbf{R}_{iz}}\right)^{v} \\
& \times \exp(2\pi i \mathbf{m} \cdot \mathbf{R}_i). \qquad (3.5.3.12)
\end{aligned}$$

$E_{\text{self}}(\varrho_i)$ is given by

$$\begin{aligned}
E_{\text{self}}(\varrho_i) = {} & \lim_{\mathbf{R} \to \mathbf{0}} \sum_{t_1 u_1 v_1} \sum_{t_2 u_2 v_2} (-1)^{(t_2+u_2+v_2)} \\
& \times \left(\frac{\partial}{\partial \mathbf{R}_\mathbf{x}}\right)^{t_1+t_2} \left(\frac{\partial}{\partial \mathbf{R}_\mathbf{y}}\right)^{u_1+u_2} \left(\frac{\partial}{\partial \mathbf{R}_\mathbf{z}}\right)^{v_1+v_2} \\
& \times \left\{ \sum_{(l_1,l_2) \in C \times C} \mathbf{c}_{i,l_1,t_1 u_1 v_1} \mathbf{c}_{i,l_2,t_2 u_2 v_2} \frac{\operatorname{erf}(\beta^{1/2}|\mathbf{R}|)}{|\mathbf{R}|} \right. \\
& \left. + \sum_{(l_1,l_2) \notin C \times C} \mathbf{c}_{i,l_1,t_1 u_1 v_1} \mathbf{c}_{i,l_2,t_2 u_2 v_2} \frac{\operatorname{erf}(\mu_{12}^{1/2}|\mathbf{R}|)}{|\mathbf{R}|} \right\},
\end{aligned}$$

where $1/\mu_{12} = 1/\theta_{l_1} + 1/\theta_{l_2}$, and the unit-cell dipole components are given by

$$\begin{aligned}
\mathbf{D}_x &= \sum_{i=1}^{N} q_i \mathbf{R}_x + \sum_{i=1}^{N} \sum_{l=1}^{L} \mathbf{c}_{i,l,100}, \\
\mathbf{D}_y &= \sum_{i=1}^{N} q_i \mathbf{R}_y + \sum_{i=1}^{N} \sum_{l=1}^{L} \mathbf{c}_{i,l,010}, \\
\mathbf{D}_z &= \sum_{i=1}^{N} q_i \mathbf{R}_z + \sum_{i=1}^{N} \sum_{l=1}^{L} \mathbf{c}_{i,l,001}.
\end{aligned}$$

In the limit that all Gaussians are compact with $\theta_l \to \infty$, $E_{S,K}(\underline{\varrho}^{\{\mathbf{N}\}})$ reduces to the Coulomb energy of ideal multipoles positioned at the expansion points.

3.5.4. Computational efficiency

The Ewald-sum methods discussed above lead to calculations that are much faster than straightforward summation, as well as clarifying the nature of the conditional convergence. However, there are opportunities for further significant speed-ups, which we cover in this section. The computational cost of the Ewald sum is dominated by the direct and reciprocal sum. We cover these in turn.

3.5.4.1. The direct sum

For all the cases discussed above – point charge, normalized spherical Gaussians and the more complex densities made up of sums over a basis of normalized Hermite Gaussians – the Ewald direct sum involves terms like $|\mathbf{R}_{12} - \mathbf{n}|$ and in the Hermite case partial derivatives with respect to the x, y and z components of these terms. Given the efficiency opportunities in the reciprocal sum to be discussed below, it is advantageous to choose an Ewald parameter (referred to above as α in the point-charge case or β for the compact–compact Gaussian or Hermite Gaussian interactions) so that direct-sum terms can be neglected past a fixed cutoff, *i.e.* when $|\mathbf{R}_{12} - \mathbf{n}| > R_C$. Given a fixed Ewald parameter α, the cutoff may be fixed independently of the unit-cell size. At constant or near constant material density (number of atoms per cubic ångstrom) this leads to 'linear scaling', *i.e.* the computational cost of the Ewald direct sum grows linearly with the unit-cell volume or number of atoms in the unit cell U.

For the point-charge or normalized spherical Gaussian case, the remaining issue is efficient evaluation of the erfc function. Various table lookup schemes have been advanced, as well as a polynomial expression (Allen & Tildesley, 1987). Ewald summation for ideal multipoles up to quadrupole have been discussed by Smith (1982, 1998) and Aguado & Madden (2003), using optimized combinations of terms to minimize the cost of the direct sum. A more general scheme, covering higher-order multipoles, was proposed by Sagui *et al.* (2004). In the latter

scheme the McMurchie–Davidson recursion (McMurchie & Davidson, 1978; Helgaker *et al.*, 2000) was used to calculate the higher-order derivatives of $\mathrm{erfc}(\alpha^{1/2}r)/r$ for the multipole–multipole interaction tensor. This same approach was used by Cisneros *et al.* (2006) to calculate the interaction tensor for normalized Hermite Gaussian interactions. Although originally formulated for Hermite Gaussian interactions, it applies to more general partial derivative calculations. Let $\mathbf{r} = (x, y, z)$ be a point in $\Re^3$ with $|\mathbf{r}| = r$, and let $g(r)$ be a smooth function of r. For example, $g(r) = \mathrm{erf}(\alpha^{1/2}r)/r$. We are interested in the partial derivatives with respect to x, y and z of $g(r)$. Let $R(0, 0, 0, 0) = g(r)$, and for $n > 0$ let $R(0, 0, 0, n) = (1/r)(\mathrm{d}/\mathrm{d}r)R(0, 0, 0, n-1)$. For non-negative integers t, u, v, let $R(t, u, v, n) = (\partial/\partial x)^t(\partial/\partial y)^u(\partial/\partial z)^v R(0, 0, 0, n)$. Then we have the following recursion:

$$
\begin{aligned}
R(t+1, u, v, n) &= xR(t, u, v, n+1) + tR(t-1, u, v, n+1),\\
R(t, u+1, v, n) &= yR(t, u, v, n+1) + uR(t, u-1, v, n+1),\\
R(t, u, v+1, n) &= zR(t, u, v, n+1) + vR(t, u, v-1, n+1).
\end{aligned}
$$

The proof of the last recursion depends on the fact that $\partial/\partial z = (z/r)\partial/\partial r$ and on Leibniz' generalization of the product rule for differentiation, namely $(fg)^{(k)} = \sum_{j=0}^{k} \binom{k}{j} f^{(j)} g^{(k-j)}$, and is as follows:

$$
\begin{aligned}
R(t, u, v+1, n) &= (\partial/\partial x)^t(\partial/\partial y)^u(\partial/\partial z)^v(\partial/\partial z)R(0, 0, 0, n)\\
&= (\partial/\partial x)^t(\partial/\partial y)^u(\partial/\partial z)^v z(1/r)\\
&\quad \times (\mathrm{d}/\mathrm{d}r)R(0, 0, 0, n)\\
&= (\partial/\partial x)^t(\partial/\partial y)^u(\partial/\partial z)^v zR(0, 0, 0, n+1)\\
&= (\partial/\partial x)^t(\partial/\partial y)^u[z(\partial/\partial z)^v R(0, 0, 0, n+1)\\
&\quad + v(\partial/\partial z)^{v-1}R(0, 0, 0, n+1)]\\
&= zR(t, u, v, n+1) + vR(t, u, v-1, n+1).
\end{aligned}
$$

The other two recursions are proved similarly. To apply this to the case $g(r) = \mathrm{erf}(\alpha^{1/2}r)/r$ we recall from equation (3.5.3.4) that $(2/\pi^{1/2})F_0(x^2) = \mathrm{erf}(|x|)/|x|$ and that for the higher-order Boy's functions $F_n(x)$, $(\mathrm{d}/\mathrm{d}x)F_n(x) = -F_{n+1}(x)$. Efficient evaluation of the Boy's functions is discussed by Helgaker *et al.* (2000).

Further economies for point multipoles can be achieved by using symmetry. For example, it is only necessary to keep track of six quadrupolar terms, with a similar reduction for Hermite Gaussians. A general discussion, including rotation of multipoles or Hermite Gaussians from local molecule frame to laboratory frame, is provided in Sagui *et al.* (2004) as well as in Cisneros *et al.* (2006).

3.5.4.2. Reciprocal-sum speed-ups: the smooth particle mesh Ewald and fast Fourier Poisson methods

The smooth particle mesh Ewald (PME) (Essmann *et al.*, 1995) and fast Fourier Poisson (FFP) (York & Yang, 1994) methods both rely on the three-dimensional fast Fourier transform (3DFFT) for computational efficiency. In combination with the above use of cutoffs in the direct sum, they lower the computational cost of the Ewald sum from $O(N^{3/2})$ (Perram *et al.*, 1988) to $O[N\log(N)]$, where N is the number of particles and *e.g.* $O(N^{3/2})$ denotes a quantity such that $O(N^{3/2})/N^{3/2}$ remains bounded as $N \to \infty$. For the point-charge case, the key insight is that if the point charges happen to be located on a regular grid, the structure factors [equation (3.5.2.14)] become discrete Fourier transforms and can be rapidly calculated using FFTs.

Like other variants of the original particle–particle particle–mesh (P^3M) method (Hockney & Eastwood, 1981), the smooth PME method interpolates charge onto grid positions. The smooth PME, like the original PME algorithm (Darden *et al.*, 1993), formulates this problem as one of interpolation of the complex exponentials appearing in the structure factor. This interpolation process is very accurate at low frequency, *i.e.* for small $|\mathbf{m}|$, and becomes progressively less accurate as $|\mathbf{m}|$ grows. However, the weight function $\exp(-\pi^2\mathbf{m}^2/\alpha)/\mathbf{m}^2$ multiplying the structure factor in the reciprocal sum severely damps the high-frequency (large $|\mathbf{m}|$) terms, thus preserving overall accuracy. In contrast, the FFP method relies on the fact that $\exp(-\pi^2\mathbf{m}^2/\alpha)/\mathbf{m}^2 S(\mathbf{m})$ is the Fourier transform of a superposition of normalized spherical Gaussian charge distributions $q_j\rho_j$, centred at $\mathbf{r}_j$, as discussed above. This Fourier transform is approximated by the FFT of the gridded charge density obtained by direct sampling of the Gaussians onto the grid (similar to the use of the FFT in macromolecular X-ray crystallography). Note that the smooth PME method has been easily extended to the treatment of London dispersion interactions, since these involve similar structure factors with a different weight function g_p, $p = 6, 8, 10$. The extension is not so straightforward with the FFP approach. However, as discussed below, the FFP approach has advantages in the case of continuous charge density when more than one Gaussian exponent is involved.

For the case of point charges, P^3M methods based on least-squares approximation of the exact Ewald reciprocal sum gradient (Pollock & Glosli, 1996; Darden *et al.*, 1997) can be made more efficient than the smooth PME method at the same accuracy, despite requiring more FFTs (Deserno & Holm, 1998). However, for the case of ideal multipoles or normalized Hermite Gaussians, the number of such FFTs becomes prohibitive, whereas the smooth PME, which relies on analytic differentiation of B-splines, does not need more FFTs as the order of differentiation ($t + u + v$) in equation (3.5.3.12) grows.

3.5.4.2.1. The smooth PME approach

We begin by re-expressing the complex exponentials appearing in the structure factors in equations (3.5.2.14), (3.5.2.26) and (3.5.3.12). Recall the form of $\mathbf{m}$ from equation (3.5.2.2), and let M_1, M_2, M_3 be positive integers of the scale of the unit-cell dimensions. Then for $\mathbf{r} \in U$ we can write

$$
\begin{aligned}
\mathbf{m}\cdot\mathbf{r} &= m_1\mathbf{a}_1^*\cdot\mathbf{r} + m_2\mathbf{a}_2^*\cdot\mathbf{r} + m_3\mathbf{a}_3^*\cdot\mathbf{r}\\
&= z_1w_1 + z_2w_2 + z_3w_3,
\end{aligned}
$$

where $w_\alpha = M_\alpha\mathbf{a}_\alpha^*\cdot\mathbf{r}$ and $z_\alpha = m_\alpha/M_\alpha$, $\alpha = 1, 2, 3$, and thus

$$
\exp(2\pi i\mathbf{m}\cdot\mathbf{r}) = \exp(2\pi i z_1w_1)\exp(2\pi i z_2w_2)\exp(2\pi i z_3w_3).
$$

To motivate the smooth PME approach, suppose we can approximate $\exp(2\pi i z w)$ by some function $g(w; z)$ of the form

$$
g(w; z) = \lambda(z)\sum_{k=-\infty}^{\infty} B(w-k)\exp(2\pi i z k), \tag{3.5.4.1}
$$

where $B(w)$ is a well behaved function with continuous derivatives as needed for the general structure factor in equation (3.5.3.12). We first examine the structure factor for point-charge Ewald sums, defined in equation (3.5.2.14). This can be approximated as

$$
\begin{aligned}
S(\mathbf{m}) &= \sum_{j=1}^{N} q_j \exp(2\pi i z_1 w_{1j}) \exp(2\pi i z_2 w_{2j}) \exp(2\pi i z_3 w_{3j}) \\
&\simeq \lambda(z_1)\lambda(z_2)\lambda(z_3) \\
&\quad \times \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \sum_{k_3=-\infty}^{\infty} \sum_{j=1}^{N} q_j B(w_{1j}-k_1) B(w_{2j}-k_2) \\
&\quad \times B(w_{3j}-k_3) \exp[2\pi i (z_1 k_1 + z_2 k_2 + z_3 k_3)] \\
&= \lambda(z_1)\lambda(z_2)\lambda(z_3) \\
&\quad \times \sum_{k_1=-(M_1/2)+1}^{M_1/2} \sum_{k_2=-(M_2/2)+1}^{M_2/2} \sum_{k_3=-(M_3/2)+1}^{M_3/2} Q(k_1,k_2,k_3) \\
&\quad \times \exp\left[2\pi i \left(\frac{m_1 k_1}{M_1} + \frac{m_2 k_2}{M_2} + \frac{m_3 k_3}{M_3}\right)\right], \qquad (3.5.4.2)
\end{aligned}
$$

where $Q(k_1, k_2, k_3)$ is given by

$$
\begin{aligned}
Q(k_1,k_2,k_3) &= \sum_{j=1}^{N} q_j \sum_{l_1=-\infty}^{\infty} B(w_{1j}-k_1-l_1M_1) \\
&\quad \times \sum_{l_2=-\infty}^{\infty} B(w_{2j}-k_2-l_2M_2) \\
&\quad \times \sum_{l_3=-\infty}^{\infty} B(w_{3j}-k_3-l_3M_3) \qquad (3.5.4.3)
\end{aligned}
$$

and we have used the fact that $\exp[2\pi i m_\alpha (k_\alpha + M_\alpha)/M_\alpha] = \exp(2\pi i m_\alpha k_\alpha / M_\alpha)$ for $\alpha = 1, 2, 3$.

Next we turn to the general structure factors $S_l(\mathbf{m})$ from equation (3.5.3.12). Define $w_{\alpha j}, \alpha = 1, 2, 3$ as above for expansion points $\mathbf{R}_j \in U$, and note that

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{R}_{jx}} &= \frac{\partial}{\partial w_{1j}} \frac{\partial w_{1j}}{\partial \mathbf{R}_{jx}} + \frac{\partial}{\partial w_{2j}} \frac{\partial w_{2j}}{\partial \mathbf{R}_{jx}} + \frac{\partial}{\partial w_{3j}} \frac{\partial w_{3j}}{\partial \mathbf{R}_{jx}} \\
&= b_1 \frac{\partial}{\partial w_{1j}} + b_2 \frac{\partial}{\partial w_{2j}} + b_3 \frac{\partial}{\partial w_{3j}},
\end{aligned}
$$

where b_1, b_2, b_3 are constants depending on the reciprocal bases $\mathbf{a}_\alpha^*$ and on M_α, $\alpha = 1, 2, 3$. Similar expansions hold for the other partial derivatives. Thus we can approximate the general structure factor $S_l(\mathbf{m})$ by

$$
\begin{aligned}
S_l(\mathbf{m}) &= \sum_{j=1}^{N} \sum_{tuv} \mathbf{c}_{j,l,tuv} \left(\frac{\partial}{\partial \mathbf{R}_{jx}}\right)^t \left(\frac{\partial}{\partial \mathbf{R}_{jy}}\right)^u \left(\frac{\partial}{\partial \mathbf{R}_{jz}}\right)^v \\
&\quad \times \exp(2\pi i \mathbf{m} \cdot \mathbf{R}_j) \\
&= \sum_{j=1}^{N} \sum_{tuv} \mathbf{d}_{j,l,tuv} \left(\frac{\partial}{\partial w_{1j}}\right)^t \left(\frac{\partial}{\partial w_{2j}}\right)^u \left(\frac{\partial}{\partial w_{3j}}\right)^v \\
&\quad \times \exp(2\pi i z_1 w_{1j}) \exp(2\pi i z_2 w_{2j}) \exp(2\pi i z_3 w_{3j}) \\
&\simeq \lambda(z_1)\lambda(z_2)\lambda(z_3) \\
&\quad \times \sum_{k_1=-(M_1/2)+1}^{M_1/2} \sum_{k_2=-(M_2/2)+1}^{M_2/2} \sum_{k_3=-(M_3/2)+1}^{M_3/2} Q_l(k_1,k_2,k_3) \\
&\quad \times \exp\left[2\pi i \left(\frac{m_1 k_1}{M_1} + \frac{m_2 k_2}{M_2} + \frac{m_3 k_3}{M_3}\right)\right], \qquad (3.5.4.4)
\end{aligned}
$$

where $\mathbf{d}_{j,l,tuv}$ are the transformed Hermite coefficients obtained from $\mathbf{c}_{j,l,tuv}$ by change of variables, using b_1, b_2, b_3 and where $Q_l(k_1, k_2, k_3)$ is given by

$$
\begin{aligned}
Q_l(k_1,k_2,k_3) &= \sum_{j=1}^{N} \sum_{tuv} \mathbf{d}_{j,l,tuv} \\
&\quad \times \sum_{l_1=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{1j}}\right)^t B(w_{1j}-k_1-l_1M_1) \\
&\quad \times \sum_{l_2=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{2j}}\right)^t B(w_{2j}-k_2-l_2M_2) \\
&\quad \times \sum_{l_3=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{3j}}\right)^t B(w_{3j}-k_3-l_3M_3).
\end{aligned}
$$

In equation (3.5.4.2) we recognize the approximation to $S(\mathbf{m})$ as $\lambda(z_1)\lambda(z_2)\lambda(z_3)$ times the three-dimensional discrete Fourier transform of $Q(k_1, k_2, k_3)$, which can be calculated rapidly by the 3DFFT. If the function $B(w)$ has finite support then $Q(k_1, k_2, k_3)$ can be calculated rapidly [in order(N) time] as well. Similarly we have the 3DFFT of $Q_l(k_1, k_2, k_3)$ in equation (3.5.4.4). Thus the efficient calculation of structure factors depends on the approximation given in equation (3.5.4.1), which in turn rests on the choice of $B(w)$. The smooth PME relies on B-splines.

3.5.4.2.1.1. Properties of B-splines

We proceed by deriving the facts needed about B-splines. The B-splines of order n, $B_n(w)$ are functions of a real variable defined as follows:

$$
B_1(w) = \begin{cases} 1 & \text{if } 0 \le w \le 1 \\ 0 & \text{otherwise} \end{cases}
$$

$$
B_2(w) = \begin{cases} w & \text{if } 0 \le w \le 1 \\ 2-w & \text{if } 1 \le w \le 2 \\ 0 & \text{otherwise} \end{cases}
$$

$$
\begin{aligned}
B_{n+1}(w) &= \int_{-\infty}^{\infty} B_n(w-v) B_1(v)\, \mathrm{d}v \\
&= \int_{0}^{1} B_n(w-v) B_1(v)\, \mathrm{d}v.
\end{aligned}
$$

We use the following properties of the B-splines:

(1) for $n \ge 1$ and all w, $0 \le B_n(w)$ and $B_n(w) = 0$ for $w < 0$ or $w > n$;

(2) for $n \ge 1$ and all w, $\sum_{j=-\infty}^{\infty} B_n(w-j) = 1$;

(3) for $n \ge 1$, $0 \le w \le n$, $B_n(n-w) = B_n(w)$;

(4) for $n \ge 2$, $0 \le w \le n$, $(\mathrm{d}/\mathrm{d}w) B_n(w) = B_{n-1}(w) - B_{n-1}(w-1)$;

(5) for $n \ge 2$, $0 \le w \le n$, $B_n(w) = [wB_{n-1}(w) + (n-w)B_{n-1}(w-1)]/(n-1)$.

The first property to be established is clearly true for $n = 1$ and if true for n, then in the above integral recursively defining B_{n+1} if $w < 0$ or $w > (n+1)$ the integrand must be zero for $0 \le v \le 1$, so $B_{n+1}(w) = 0$. Thus it is true in general by mathematical induction.

The second property is also clearly true for $n = 1$. Assume it is true for n. Then $\sum_{j=-\infty}^{\infty} B_{n+1}(w-j) = \int_0^1 [\sum_{j=-\infty}^{\infty} B_n(w-v-j)] B_1(v)\, \mathrm{d}v = \int_0^1 B_1(v)\, \mathrm{d}v = 1$. Thus it is true in general by mathematical induction.

Next we show symmetry: $B_n(n-w) = B_n(w)$, $0 \le w \le n$. Again this is clearly true for $n = 1$. Assuming it is true for n, we have

$$\begin{aligned}
B_{n+1}(n+1-w) &= \int_0^1 B_n(n+1-w-v)B_1(v)\,\mathrm{d}v \\
&= \int_0^1 B_n[n+1-w-(1-v')]B_1(1-v')\,\mathrm{d}v' \\
&= \int_0^1 B_n[n-(w-v')]B_1(1-v')\,\mathrm{d}v' \\
&= \int_0^1 B_n(w-v')B_1(v')\,\mathrm{d}v' \\
&= B_{n+1}(w),
\end{aligned}$$

where we substituted $v = (1 - v')$ in the second line above.

The fourth property is an efficient recursion for the derivatives of B-splines. It can be seen directly for $n = 2$ for $w \neq 0, 1$, or 2. Assume it is true for $n \geq 2$. Then for $0 \leq w \leq n + 1$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}w}B_{n+1}(w) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}w}B_n(w-v)B_1(v)\,\mathrm{d}v \\
&= \int_0^1 B_{n-1}(w-v)B_1(v)\,\mathrm{d}v \\
&\quad - \int_0^1 B_{n-1}(w-1-v)B_1(v)\,\mathrm{d}v \\
&= B_n(w) - B_n(w-1),
\end{aligned}$$

thus it is true in general by induction.

The final property is an efficient recursion for $B_n(w)$. It can be seen directly for $n = 2$. Assume it is true for $n \geq 2$. Then for $0 \leq w \leq n + 1$,

$$\begin{aligned}
B_{n+1}(w) &= \int_0^1 \left[\frac{w-v}{n-1}B_{n-1}(w-v)\right. \\
&\quad \left. + \frac{n-(w-v)}{n-1}B_{n-1}(w-v-1)\right]B_1(v)\,\mathrm{d}v \\
&= \frac{n}{n-1}B_n(w-1) \\
&\quad + \frac{1}{n-1}\int_0^1 (w-v)\frac{\mathrm{d}}{\mathrm{d}w}B_n(w-v)B_1(v)\,\mathrm{d}v \\
&= \frac{n}{n-1}B_n(w-1) \\
&\quad - \frac{1}{n-1}\int_0^1 (w-v)\frac{\mathrm{d}}{\mathrm{d}v}B_n(w-v)B_1(v)\,\mathrm{d}v \\
&= \frac{n}{n-1}B_n(w-1) \\
&\quad - \frac{1}{n-1}[(w-1)B_n(w-1) - wB_n(w)] \\
&\quad + \frac{1}{n-1}\int_0^1 B_n(w-v)\frac{\mathrm{d}}{\mathrm{d}v}(w-v)B_1(v)\,\mathrm{d}v \\
&= \frac{n+1-w}{n-1}B_n(w-1) + \frac{w}{n-1}B_n(w) \\
&\quad - \frac{1}{n-1}B_{n+1}(w),
\end{aligned}$$

where we have used property (4) in the third line and integration by parts in the fifth and sixth lines. The induction step follows by moving the last term on the right-hand side over to the left side of the equation.

3.5.4.2.1.2. B-spline approximation to trigonometric functions

Next we examine properties of the B-spline approximations to the complex exponentials that appear in the structure factors. Let $g_n(w; z)$ be given by

$$g_n(w; z) = \lambda_n(z) \sum_{j=-\infty}^{\infty} B_n(w-j)\exp(2\pi i z j). \tag{3.5.4.5}$$

The complex number $\lambda_n(z)$ is chosen to optimize the approximation of $\exp(2\pi i z w)$ by $g_n(w; z)$ as shown below. It is convenient to first study the properties of the function $h_n(w; z)$ given by

$$\begin{aligned}
h_n(w; z) &= \exp(-2\pi i z w)g_n(u; z) \\
&= \lambda_n(z) \sum_{j=-\infty}^{\infty} B_n(w-j)\exp[-2\pi i z(w-j)],
\end{aligned}$$

which is an approximation to the constant one. From its definition it is clear that $h_n(w; z)$ is periodic in w with period one, *i.e.* $h_n(w + 1; z) = h_n(w; z)$. Therefore it can be expanded in a Fourier series:

$$h_n(w; z) = \sum_{k=-\infty}^{\infty} c_n(k; z)\exp(2\pi i k w). \tag{3.5.4.6}$$

The Fourier coefficients are given by

$$\begin{aligned}
c_n(k; z) &= \int_0^1 h_n(w; z)\exp(-2\pi i k w)\,\mathrm{d}w \\
&= \lambda_n(z)\int_0^1 \sum_{j=-\infty}^{\infty} B_n(w-j) \\
&\quad \times \exp[-2\pi i z(w-j)]\exp(-2\pi i k w)\,\mathrm{d}u \\
&= \lambda_n(z) \sum_{j=-\infty}^{\infty}\int_0^1 B_n(w-j) \\
&\quad \times \exp[-2\pi i(z+k)(w-j)]\,\mathrm{d}u \\
&= \lambda_n(z) \int_{-\infty}^{\infty} B_n(v)\exp[-2\pi i(z+k)v]\,\mathrm{d}v \\
&= \lambda_n(z)\ \hat{B}_n(z+k),
\end{aligned}$$

where $\hat{B}_n$ denotes the Fourier transform of B_n. Since B_n are defined recursively by convolutions, its Fourier transform is obtained simply from that of B_1: $\hat{B}_n = (\hat{B}_1)^n$. Here we are using the fact that if f, g and h are related by convolution, $h(w) = (f \star g)(w) = \int_{-\infty}^{\infty} f(w-v)g(v)\,\mathrm{d}v$, then $\hat{h} = \hat{f}\hat{g}$. Since $\hat{B}_1$ can be obtained by straightforward integration, we have

$$c_n(k; z) = \begin{cases} \lambda_n(z) & \text{if } z + k = 0 \\ \lambda_n(z)\left[\dfrac{1-\exp(-2\pi i z)}{2\pi i}\right]^n \left(\dfrac{1}{z+k}\right)^n & \text{otherwise.} \end{cases}$$

The original smooth PME method set $\lambda_n(z) = \bar{\lambda}_n(z)$, where $\bar{\lambda}_n(z)$ was chosen to enforce interpolation: $g_n(w; z) = \exp(2\pi i z w)$ or $h_n(w; z) = 1$ for integer values of w. This was related to the so-called Euler spline, studied for more general complex-valued interpolation problems by Schoenberg (1973). The special case of interpolating the function $\exp(2\pi i z w)$ is simpler, due to the Fourier-series representation, allowing us to to derive necessary properties in a self-contained way. From equations (3.5.4.6) and the above expression for $c_n(k; z)$, noting that $c_n(k; z) = 0$ for

integral z with $z + k \neq 0$, the interpolation condition can be written

$$1 = \begin{cases} \bar{\lambda}_n(z) & z \text{ an integer} \\ \bar{\lambda}_n(z)\left[\dfrac{1-\exp(-2\pi i z)}{2\pi i}\right]^n \displaystyle\sum_{k=-\infty}^{\infty}\left(\frac{1}{z+k}\right)^n & \text{otherwise.} \end{cases}$$

Note that except for $\bar{\lambda}_n(z)$, the expressions appearing in the interpolation condition are periodic in z with period one. The interpolation condition thus forces $\bar{\lambda}_n(z)$ to be periodic as well, and so we focus attention on the range $-\frac{1}{2} \leq z \leq \frac{1}{2}$. If $n \geq 2$ the sum in the above interpolation condition is bounded, so in order to be able to choose $\bar{\lambda}_n(z)$ to satisfy the interpolation condition it is sufficient to demonstrate that $\sum_{k=-\infty}^{\infty}[1/(z+k)]^n \neq 0$ for $-\frac{1}{2} \leq z \leq \frac{1}{2}$. If n is even, the terms in this sum are all positive so it is always possible to interpolate. If n is odd and $z = \frac{1}{2}$ then the sum is zero, since $(1/z)^n + [1/(-1+z)]^n = 0$, $[1/(1+z)^n + 1/(-2+z)]^n = 0$ *etc.* Similarly the interpolation condition fails for $z = -\frac{1}{2}$. However, for $0 < z < \frac{1}{2}$, $(1/z)^n + [1/(-1+z)]^n > 0$, $[1/(1+z)^n + 1/(-2+z)]^n > 0$ *etc.* Similarly for $-\frac{1}{2} < z < 0$. Thus we can say that for $n \geq 2$ and $-\frac{1}{2} < z < \frac{1}{2}$ the interpolation condition holds, and in this case we can write h_n as

$$h_n(w; z) = \sum_{k=-\infty}^{\infty} \gamma_n(k; z)\exp(2\pi i k w), \tag{3.5.4.7}$$

where

$$\gamma_n(k; z) = \left(\frac{1}{z+k}\right)^n \Bigg/ \sum_{j=-\infty}^{\infty}\left(\frac{1}{z+j}\right)^n. \tag{3.5.4.8}$$

The interpolating $g_n(w; z)$ is obtained from this h_n by multiplying by $\exp(2\pi i z w)$. Numerically, it is more convenient to use the alternative expression from equation (3.5.4.5), with $\bar{\lambda}_n(z)$ given by

$$\bar{\lambda}_n(z) = \left[\sum_{j=0}^{n} B_n(j)\exp(-2\pi i z j)\right]^{-1}.$$

Note that due to the above results for h_n, $\sum_{j=0}^{n} B_n(j)\exp(-2\pi i z j) \neq 0$ if $n \geq 2$ and $-\frac{1}{2} < z < \frac{1}{2}$.

Subsequent to the work by Essmann *et al.* (1995), we found (Darden *et al.*, 1997) that a more accurate approximation was given by least-squares fitting of $\exp(2\pi i z w)$ by $g_n(w; z)$, or equivalently 1 by $h_n(w; z)$. Thus, with h_n given by equation (3.5.4.7) we are looking for a complex number $\tilde{\lambda} = \tilde{\lambda}_n(z)$ that minimizes the mean-square error MSE, given by

$$\begin{aligned} \text{MSE} &= \int_0^1 [1 - \tilde{\lambda}h_n(w; z)][1 - \tilde{\lambda}h_n(w; z)]^* \, \mathrm{d}w \\ &= 1 - 2Re(\tilde{\lambda})\gamma_n(0; z) + [Re(\tilde{\lambda})^2 + Im(\tilde{\lambda})^2]\sum_{k=-\infty}^{\infty}\gamma_n^2(k; z), \end{aligned}$$

where $Re(\tilde{\lambda})$ and $Im(\tilde{\lambda})$ denote the real and imaginary parts of $\tilde{\lambda}$. Since $\sum_{k=-\infty}^{\infty}\gamma_n^2(k; z) > 0$, it is clear that we can set $Im(\tilde{\lambda})$ to zero, that is, restrict ourselves to real $\tilde{\lambda}$. Then, minimizing over real λ we have

$$\tilde{\lambda}_n(z) = \gamma_n(0; z) \Bigg/ \sum_{k=-\infty}^{\infty}\gamma_n^2(k; z),$$

with $\gamma_n(k; z)$ given in equation (3.5.4.8).

Thus we arrive at the least-squares optimal approximation to $\exp(2\pi i z u)$ by setting $\lambda_n(z) = \tilde{\lambda}_n(z)\bar{\lambda}_n(z)$ in equation (3.5.4.5). The mean-square error due to this can be seen to be

$$\begin{aligned} \text{MSE} &= 1 - \gamma_n^2(0; z) \Bigg/ \sum_{k=-\infty}^{\infty}\gamma_n^2(k; z) \\ &= \sum_{k\neq 0}\gamma_n^2(k; z) \Bigg/ \sum_{k=-\infty}^{\infty}\gamma_n^2(k; z) \end{aligned}$$

and thus the error in the approximation is seen to be small for small $|z|$ but increasing as $|z| \to \frac{1}{2}$, as discussed above (the reciprocal Ewald weight term compensates for the increased error by rapidly damping the larger $|z|$ terms). For $0 < |z| < \frac{1}{2}$ the error decreases to zero as the B-spline order $n \to \infty$. The accuracy of the PME is thus increased by raising the order of interpolation and/or the grid size M_1, M_2, M_3, which lowers $|z|$ for a given $\mathbf{m}$. When derivatives with respect to w of $g_n(w; z)$ are considered in this latter representation, their effect is to effectively lower the order of interpolation in the coefficients γ_n. Thus the order of interpolation needs to be higher, the higher the order of differentiation considered.

3.5.4.2.2. The fast Fourier Poisson approach

The fast Fourier Poisson (FFP) method (York & Yang, 1994) accelerates the calculation of the structure factors in equations (3.5.2.14) and (3.5.3.12) in much the same way as FFT methods accelerate structure-factor and density-map calculations in macromolecular structure determination. Originally the method was used to calculate structure factors appearing in Ewald sums over point charges. It is easier to see the basic idea if we rewrite the Ewald reciprocal sum from equation (3.5.2.15) as follows:

$$\begin{aligned} E_r(\underline{\mathbf{r}}^{\{\mathbf{N}\}}) &= \frac{1}{2\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{\exp(-\pi^2\mathbf{m}^2/\alpha)}{\mathbf{m}^2}S(\mathbf{m})S(-\mathbf{m}) \\ &= \frac{1}{2\pi V}\sum_{\mathbf{m}\neq\mathbf{0}}\frac{1}{\mathbf{m}^2}\tilde{S}(\mathbf{m})\tilde{S}(-\mathbf{m}), \end{aligned}$$

where the modified structure factor $\tilde{S}(\mathbf{m})$ is given by

$$\tilde{S}(\mathbf{m}) = \exp(-\pi^2\mathbf{m}^2/2\alpha)\sum_{j=1}^{N} q_j \exp(2\pi i \mathbf{m}\cdot\mathbf{r}_j).$$

We recognize $\tilde{S}(\mathbf{m})$ as the three-dimensional Fourier transform, evaluated at $\mathbf{m}$, of a density $\tilde{\rho}(\mathbf{r})$ given by

$$\tilde{\rho}(\mathbf{r}) = \sum_{j=1}^{N} q_j \rho_{2\alpha}(\mathbf{r} - \mathbf{r}_j),$$

where $\rho_{2\alpha}$ is a normalized Gaussian with Gaussian exponent 2α.

In turn we approximate this integral by a discrete Fourier transform. As above let M_1, M_2, M_3 be positive integers of the scale of the unit-cell dimensions. Let $\delta v = V/(M_1M_2M_3)$, where V is the volume of the unit cell. Recall the lattice basis vectors $\mathbf{a}_\alpha$, $\alpha = 1, 2, 3$. Given integers k_1, k_2, k_3 define $\tilde{\rho}_0(k_1, k_2, k_3)$ by $\tilde{\rho}_0(k_1, k_2, k_3) = \tilde{\rho}[(k_1/M_1)\mathbf{a}_1 + (k_2/M_2)\mathbf{a}_2 + (k_3/M_3)\mathbf{a}_3]$. Then we can approximate the modified structure factor $\tilde{S}(\mathbf{m})$ by

$$\tilde{S}(\mathbf{m}) \simeq \sum_{k_1=-(M_1/2)+1}^{M_1/2} \sum_{k_2=-(M_2/2)+1}^{M_2/2} \sum_{k_3=-(M_3/2)+1}^{M_3/2} Q(k_1,k_2,k_3) \times \exp\left[2\pi i\left(\frac{m_1k_1}{M_1}+\frac{m_2k_2}{M_2}+\frac{m_3k_3}{M_3}\right)\right],$$

where $Q(k_1,k_2,k_3)$ is given by

$$Q(k_1,k_2,k_3) = \sum_{l_1,l_2,l_3=-\infty}^{\infty} \tilde{\rho}_0(k_1+l_1M_1, k_2+l_2M_2, k_3+l_3M_3)\cdot \delta v.$$

The FFP can be generalized to Hermite Gaussians by sampling that density in place of the superposition of spherical Gaussians. In equation (3.5.3.11) the term

$$(1/2\pi V)\sum_{\mathbf{m}\neq\mathbf{0}}(1/\mathbf{m}^2)\exp(-\pi^2\mathbf{m}^2/2\beta)\sum_{l_1\in C}S_{l_1}(\mathbf{m}) \times \exp(-\pi^2\mathbf{m}^2/2\beta)\sum_{l_2\in C}S_{l_2}(-\mathbf{m})$$

can be handled by the methods just outlined. However the term

$$(1/2\pi V)\sum_{\mathbf{m}\neq\mathbf{0}}(1/\mathbf{m}^2)\sum_{(l_1,l_2)\notin C\times C}\exp(-\pi^2\mathbf{m}^2/\theta_{l_1})S_{l_1}(\mathbf{m}) \times \exp(-\pi^2\mathbf{m}^2/\theta_{l_2})S_{l_2}(-\mathbf{m})$$

can present problems. The difficulty comes in the case when, for example, l_1 is compact but l_2 is not. Then the combined Gaussian weights are sufficiently damping to efficiently converge the series, but it can be very inefficient to directly sample the superposition of Hermite Gaussians with a compact exponent. A similar problem was faced by Fusti-Molnar & Pulay (2002). Their solution was to filter out high frequencies in the Fourier-transformed compact densities and back transform them to obtain diffuse real-space densities. Instead, Cisneros *et al.* (2006) adjusted the exponents of both the compact and diffuse Gaussians. Choose a $\bar{\theta}$ such that if θ_l is diffuse $1/\theta_l > 1/\bar{\theta}$. For example, choose $\bar{\theta} = 1/(4\beta)$. As before, define $\tilde{S}_l(\mathbf{m})$ by $\tilde{S}_l(\mathbf{m}) = \exp(-\pi^2\mathbf{m}^2/\theta_l)S_l(\mathbf{m})$, and $\tilde{S}_{l+}(\mathbf{m})$, $\tilde{S}_{l-}(\mathbf{m})$ by

$$\tilde{S}_{l+}(\mathbf{M}) = \exp\left[-\pi^2\mathbf{m}^2\left(\frac{1}{\theta_l}+\frac{1}{4\beta}\right)\right]S_l(\mathbf{m})$$
$$\tilde{S}_{l-}(\mathbf{M}) = \exp\left[-\pi^2\mathbf{m}^2\left(\frac{1}{\theta_l}-\frac{1}{4\beta}\right)\right]S_l(\mathbf{m}).$$

Then, if $l_1 \in C$ and $l_2 \in D$,

$$\sum_{l_1\in C}\sum_{l_2\in D}\tilde{S}_{l_1}(\mathbf{m})\tilde{S}_{l_2}(\mathbf{m}) = \sum_{l_1\in C}\sum_{l_2\in D}\tilde{S}_{l_1+}(\mathbf{m})\tilde{S}_{l_2-}(\mathbf{m}).$$

Thus instead of approximating $\tilde{S}_l(\mathbf{m})$ by sampling the Gaussian followed by applying the 3DFFT, they approximated $\tilde{S}_{l+}(\mathbf{m})$ or $\tilde{S}_{l-}(\mathbf{m})$. Note that with the above choice of $\bar{\theta}$, for any $l \in C$, $\theta_{l+} < 4\beta$ and for any $l \in D$, $\theta_{l-} < 4\beta$. That is, all the modified Gaussians can be sampled on a uniform grid. Finally, note that for the diffuse–diffuse interactions

$$\sum_{l_1\in D}\sum_{l_2\in D}\tilde{S}_{l_1}(\mathbf{m})\tilde{S}_{l_2}(\mathbf{m}) = \sum_{l_1\in D}\sum_{l_2\in D}\tilde{S}_{l_1-}(\mathbf{m})\tilde{S}_{l_2-}(\mathbf{m})\exp(-\pi^2\mathbf{m}^2/2\beta),$$

i.e. the interaction between adjusted diffuse Gaussians can be corrected efficiently in reciprocal space.

Table 3.5.4.1. *Timing versus accuracy for smooth PME, FFP and regular Ewald sum reciprocal-space approaches*

(*a*) Time in seconds for A2 model density.

	RMS force error 10^{-3}			RMS force error 10^{-4}		
Box size	PME	FFP	Ewald	PME	FFP	Ewald
64	0.106	0.142	0.365	0.144	0.274	0.520
128	0.218	0.271	1.336	0.287	0.380	1.869
256	0.387	0.528	5.239	0.517	0.846	7.256
512	0.837	1.100	17.881	1.104	1.534	25.511
1024	1.701	2.236	71.513	2.249	3.152	108.158

(*b*) Time in seconds for P1 model density

	RMS force error 10^{-3}			RMS force error 10^{-4}		
Box size	PME	FFP	Ewald	PME	FFP	Ewald
64	0.310	0.399	2.321	0.478	0.688	3.858
128	0.591	0.832	7.706	0.923	1.576	11.056
256	1.186	1.920	35.178	1.805	2.736	49.107
512	2.549	3.825	119.863	3.794	5.684	183.487
1024	4.953	6.890	486.384	6.947	11.307	714.644

To test the efficiency of these reciprocal-space approaches, the electrostatic energy and intermolecular forces were calculated for a series of water boxes (64, 128, 256, 512 and 1024 water molecules) under periodic boundary conditions. Similar calculations were performed by Cisneros *et al.* (2006). As in that paper, the electron density of the water molecule was modelled by a five-site expansion consisting of the oxygen, the two hydrogens and the two bond midpoints, and using standard auxiliary basis functions, denoted A2 and P1, from the density-functional literature. These basis functions were obtained from the Environmental Molecular Sciences Laboratory at Pacific Northwest Laboratories (http://www.emsl.pnl.gov/forms/basisform.html). The A2 basis set expansion uses 21 Hermite primitives with 102 Hermite coefficients per water molecule, resulting in a root-mean-square relative molecular force error of about 8% compared to 6-31G* B3LYP intermolecular electrostatics (Cisneros *et al.*, 2006). The more complex P1 basis set expansion uses 38 Hermite primitives with 200 Hermite coefficients per water molecule, reducing the relative molecular force error to about 2%. For each water box the electrostatic energy and forces for the model A2 or P1 densities were calculated to a relative root-mean-square molecular force error of 10^{-3} or 10^{-4}, compared to exact Ewald summation of the respective model density.

The time to perform these calculations is given in Table 3.5.4.1. Note that the times given here are considerably faster than those reported in Figs. 13 and 14 of Cisneros *et al.* (2006). This is partly due to the faster processor used here (3.6 GHz Intel Xeon *versus* 3.2 GHz Xeon used previously; Intel compiler with similar options used in both cases) but mainly due to the more efficient algorithms described herein. Note that the PME and FFP timings increase in an approximately linear fashion as a function of box size whereas the Ewald sum timings behave quadratically. For each method the timings are a continuous function of the compact–diffuse separating exponent 2β described above in Section 3.5.3.1, except when 2β equals a Gaussian exponent in the model basis set. When this happens, a set of Hermite primitives change from compact to diffuse or *vice versa* and the relative

balance of direct-sum or reciprocal-sum calculation changes discontinuously. For the smooth PME calculations it was optimal to choose the separating exponent at one of these boundaries. This separating exponent value was then retained for the FFP and Ewald calculations, although the FFP method could be made slightly more efficient by further varying this parameter, and the Ewald summation could be considerably accelerated to an order $N^{3/2}$ method, where N is the number of water molecules, by making the separating exponent a decreasing function of system size. Neither of these optimizations would change the relative efficiencies, however.

The error in the FFP for a given Gaussian exponent is controlled by the size of the grid M_1, M_2, M_3 as in the PME method, and also by the completeness in the Gaussian sampling (the cutoff in the tails of the Gaussian). Typically, for equivalent accuracy, the FFP requires more extensive sampling onto grid points than does the smooth PME. That is, each Gaussian centred at an expansion point must be sampled over a considerably larger set of points than the tensor product B-spline in equation (3.5.4.3). The additional cost due to this has been ameliorated somewhat by the Gaussian split Ewald approach (Shan *et al.*, 2005). Additionally, in rectangular unit cells fast Poisson solvers such as multigrid (Beck, 2000; Sagui & Darden, 2001) can be applied, which in principle should parallelize more efficiently than the 3DFFT. Finally, although the smooth PME is typically more efficient for a single Gaussian exponent structure factor, the FFP approach has the strong advantage that the various Gaussian superpositions having possibly different exponents can be sampled and then fast Fourier transformed together, whereas they must be kept separate in the PME approach. Thus, as here Cisneros *et al.* (2006) found that the smooth PME approach was more efficient for smaller auxiliary basis sets, whereas they found that the FFP approach became more efficient for large auxiliary basis sets involving many different Gaussian exponents.

References

Aguado, A. & Madden, P. A. (2003). *Ewald summation of electrostatic multipole interactions up to the quadrupolar level*. *J. Chem. Phys.* **119**, 7471–7483.

Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*. Oxford: Clarendon Press.

Arfken, G. B. & Weber, H. J. (2000). *Mathematical Methods for Physicists*, 5th ed. San Diego: Academic Press.

Beck, T. L. (2000). *Real-space mesh techniques in density functional theory*. *Rev. Mod. Phys.* **72**, 1041–1080.

Böttcher, C. J. F. (1973). *Theory of Electric Polarization*. Amsterdam: Elsevier Science Publishers BV.

Broderson, S., Wilke, S., Leusen, F. J. L. & Engel, G. (2003). *A study of different approaches to the electrostatic interaction in force field methods for organic crystals*. *Phys. Chem. Chem. Phys.* **5**, 4923–4931.

Brown, D. & Neyertz, S. (1995). *A general pressure tensor calculation for molecular dynamics simulations*. *Mol. Phys.* **84**, 577–595.

Challacombe, M., White, C. & Head-Gordon, M. (1997). *Periodic boundary conditions and the fast multipole method*. *J. Chem. Phys.* **107**, 10131–10140.

Chemburkar, S. R., Bauer, J., Deming, K., Spiwek, H., Patel, K., Morris, J., Henry, R., Spanton, S., Dziki, W., Porter, W., Quick, J., Bauer, P., Donabauer, J., Narayanan, B. A., Soldani, M., Riley, D. & McFarland, K. (2000). *Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development*. *Org. Process Res. Dev.* **4**, 413–417.

Cisneros, G. A., Piquemal, J. P. & Darden, T. A. (2005). *Intermolecular electrostatics using density fitting*. *J. Chem. Phys.* **123**, 044109.

Cisneros, G. A., Piquemal, J. P. & Darden, T. A. (2006). *Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods*. *J. Chem. Phys.* **125**, 184101.

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. New York: Oxford University Press.

Coppens, P. & Volkov, A. (2004). *The interplay between experiment and theory in charge-density analysis*. *Acta Cryst.* A**60**, 357–364.

Darden, T., Pearlman, D. A. & Pedersen, L. G. (1999). *Ionic charging free energies: Spherical versus periodic boundary conditions*. *J. Chem. Phys.* **109**, 10921–10935.

Darden, T. A., Toukmaji, A. & Pedersen, L. (1997). *Long-range electrostatic effects in biomolecular simulations*. *J. Chim. Phys.* **94**, 1346–1364.

Darden, T. A., York, D. M. & Pedersen, L. G. (1993). *Particle mesh Ewald: An N log(N) method for Ewald sums in large systems*. *J. Chem. Phys.* **98**, 10089–10092.

Day, G. M., Chisholm, J., Shan, N., Motherwell, W. D. S. & Jones, W. (2004). *An assessment of lattice energy minimization for the prediction of molecular organic crystal structures*. *Cryst. Growth Des.* **4**, 1327–1340.

Day, G. M., Motherwell, W. D. S., Ammon, H. L., Boerrigter, S. X. M., Della Valle, R. G., Venuti, E., Dzyabchenko, A., Dunitz, J. D., Schweizer, B., van Eijck, B. P., Erk, P., Facelli, J. C., Bazterra, V. E., Ferrarro, M. B., Hofmann, D. W. M., Leusen, F. J. J., Liang, C., Pantelides, C. C., Karamertzanis, P. G., Price, S. L., Lewis, T. C., Nowell, H., Torrisi, A., Scheraga, H. A., Arnautova, Y. A., Schmidt, M. U. & Verwer, P. (2005). *A third blind test of crystal structure prediction*. *Acta Cryst.* B**61**, 511–527.

Day, G. M., Motherwell, W. D. S. & Jones, W. (2005). *Beyond the isotropic atom model in crystal structure prediction of rigid molecules: Atomic multipoles versus point charges*. *Cryst. Growth Des.* **5**, 1023–1033.

Deem, M. W., Newsam, J. M. & Sinha, S. K. (1990). *The h = 0 term in Coulomb sums by the Ewald transformation*. *J. Phys. Chem.* **94**, 8356–8359.

DeLeeuw, S. W., Perram, J. W. & Petersen, H. G. (1990). *Hamilton's equations for constrained dynamical systems*. *J. Stat. Phys.* **61**, 1203–1222.

Deserno, M. & Holm, C. (1998). *How to mesh up Ewald sums I: A theoretical and numerical comparison of various particle mesh routines*. *J. Chem. Phys.* **109**, 7678–7693.

Destro, R., Roversi, P., Barzaghi, M. & Marsh, R. E. (2000). *Experimental charge density of α-glycine at 23 K*. *J. Phys. Chem. A*, **104**, 1047–1054.

Dovesi, R., Orlando, R., Civalleri, B., Roetti, C., Saunders, V. R. & Zicovich-Wilson, C. M. (2005). *Crystal: a computational tool for the ab initio study of the electronic properties of crystals*. *Z. Kristallogr.* **220**, 571–573.

Dunitz, J. D. (2003). *Are crystal structures predictable? Chem. Commun.* **5**, 545–548.

Dunitz, J. D. & Bernstein, J. (1995). *Disappearing polymorphs. Acc. Chem. Res.* **28**, 193–200.

Dunitz, J. D. & Gavezzotti, A. (2005). *Molecular recognition in organic crystals: Directed intermolecular bonds or nonlocalized bonding? Angew. Chem.* **44**, 1766–1787.

Dunitz, J. D. & Scheraga, H. A. (2004). *Exercises in prognistication: Crystal structures and protein folding*. *Proc. Natl Acad. Sci. USA*, **101**, 14309–14311.

Eijck, B. P. van & Kroon, J. (1997). *Coulomb energy of polar crystals*. *J. Phys. Chem. B*, **101**, 1096–1100.

Eijck, B. P. van & Kroon, J. (2000). *Comment on 'Crystal structure prediction by global optimization as a tool for evaluating potentials: role of the dipole moment correction terms in successful predictions'*. *J. Phys. Chem. B*, **104**, 8089.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). *A smooth particle mesh Ewald method*. *J. Chem. Phys.* **103**, 8577–8593.

Eyges, L. (1980). *The Classical Electromagnetic Field*. New York: Dover Publications Inc.

Frenkel, D. & Smit, B. (2002). *Understanding Molecular Simulation*. San Diego: Academic Press.

Fusti-Molnar, L. & Pulay, P. (2002). *The Fourier transform Coulomb method: Efficient and accurate calculation of the Coulomb operator in a Gaussian basis. J. Chem. Phys.* **117**, 7827–7835.

Gavezzotti, A. (2002*a*). *Structure and intermolecular potentials in molecular crystals*. *Modelling Simul. Mater. Sci. Eng.* **10**, R1–R29.

Gavezzotti, A. (2002*b*). *Calculation of intermolecular interaction energies by direct numerical interaction energies I. Electrostatic and polarization energies in molecular crystals. J. Phys. Chem. B*, **106**, 4145–4154.

Gavezzotti, A. & Fillipini, G. (1994). *The geometry of the intermolecular x—H···y (x, y = N, O) hydrogen bond and the calibration of empirical hydrogen bond potentials*. *J. Phys. Chem.* **98**, 4831–4837.

Gill, P. M. W. & Adamson, R. D. (1996). *A family of attenuated Coulomb operators*. *Chem. Phys. Lett.* **261**, 105–110.
Helgaker, T., Jorgensen, P. & Olsen, J. (2000). *Molecular Electronic-Structure Theory*. Chichester: John Wiley and Sons.
Hockney, R. W. & Eastwood, J. W. (1981). *Computer Simulation Using Particles*. New York: McGraw-Hill.
Hummer, G. (1996). *Electrostatic potential of a homogeneously charged square and cube in two and three dimensions*. *J. Electrostat.* **3**, 285–291.
Hummer, G., Pratt, L. R. & Garcia, A. E. (1996). *On the free energy of ionic hydration*. *J. Phys. Chem.* **100**, 1206–1215.
Kittel, C. (1986). *Introduction to Solid State Physics*. New York: John Wiley and Sons.
Körner, T. W. (1988). *Fourier Analysis*. Cambridge University Press.
McMurchie, L. E. & Davidson, E. R. (1978). *One-electron and 2-electron integrals over Cartesian Gaussian functions*. *J. Comput. Phys.* **26**, 218–231.
Mooji, W. T. M., van Duijneveldt, F. B., van Duijneveldt-van de Rijdt, J. G. C. M. & van Eijck, B. P. (1999). *Transferable ab initio intermolecular potentials. 1. Derivation from methanol dimer and trimer calculations. J. Phys. Chem. A*, **103**, 9872–9882.
Nijboer, B. R. A. & Ruijgrok, Th. W. (1988). *On the energy per particle in three- and two-dimensional Wigner lattices*. *Mol. Phys.* **53**, 361–382.
Nosé, S. & Klein, M. L. (1983). *Constant pressure molecular dynamics for molecular systems*. *Mol. Phys.* **50**, 1055–1076.
Nowell, H., Frampton, C. S., Waite, J. & Price, S. L. (2006). *Blind crystal structure prediction of a novel second polymorph of 1-hydroxy-7-azabenzotriazole*. *Acta Cryst.* B**62**, 642–650.
Perram, J. W., Petersen, H. G. & DeLeeuw, S. W. (1988). *An algorithm for the simulation of condensed matter that grows as the 3/2 power of the number of particles*. *Mol. Phys.* **65**, 875–893.
Pillardy, J., Wawak, R. J., Arnautova, Y. A., Czaplewski, C. & Scheraga, H. A. (2000). *Crystal structure prediction by global optimization as a tool for evaluating potentials: Role of the dipole moment correction term in successful predictions*. *J. Am. Chem. Soc.* **122**, 907–921.
Piquemal, J. P., Cisneros, G. A. & Darden, T. A. (2006). *Towards a force field based on density fitting*. *J. Chem. Phys.* **124**, 104101.
Pollock, E. L. & Glosli, J. (1996). *Comments on p(3)m, fmm and the Ewald method for large periodic Coulombic systems*. *Comput. Phys. Comm.* **95**, 93–110.
Price, S. L. & Price, L. S. (2005). *Modelling intermolecular forces for organic crystal structure prediction*. *Struct. Bond.* **115**, 81–123.
Redlack, A. & Grindley, J. (1972). *The electrostatic potential in a finite ionic crystal*. *Can. J. Phys.* **50**, 2815–2825.
Redlack, A. & Grindley, J. (1975). *Coulombic potential lattice sums*. *J. Phys. Chem. Solids*, **36**, 73–82.
Ren, P. Y. & Ponder, J. W. (2003). *Polarizable atomic multipole water model for molecular mechanics simulation*. *J. Phys. Chem. B*, **107**, 5933–5947.
Sagui, C. & Darden, T. (2001). *Multigrid methods for classical molecular dynamics simulations of biomolecules*. *J. Chem. Phys.* **114**, 6578–6591.
Sagui, C., Pedersen, L. G. & Darden, T. A. (2004). *Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations*. *J. Chem. Phys.* **120**, 73–87.
Schoenberg, I. J. (1973). *Cardinal Spline Interpolation*. Philadelphia: Society for Industrial and Applied Mathematics.
Shan, Y. B., Klepeis, J. L., Eastwood, M. P., Dror, R. O. & Shaw, D. E. (2005). *Gaussian split Ewald: A fast Ewald mesh method for molecular simulations*. *J. Chem. Phys.* **122**, 054101.
Smith, E. R. (1981). *Electrostatic energy in ionic crystals*. *Proc. R. Soc. London Ser. A*, **373**, 27–56.
Smith, E. R. (1994). *Calculating the pressure in simulations using periodic boundary conditions*. *J. Stat. Phys.* **77**, 449–472.
Smith, W. (1982). *Point multipoles in the Ewald summation*. *CCP5 Inf. Q.* **4**, 13–25.
Smith, W. (1987). *Coping with the pressure: How to calculate the virial*. *CCP5 Inf. Q.* **26**, 43–50.
Smith, W. (1993). *Calculating the pressure*. *CCP5 Inf. Q.* **39**, 1–7.
Smith, W. (1998). *Point multipoles in the Ewald summation (revisited)*. *CCP5 Inf. Q.* **46**, 18–30.
Spackman, M. A., Weber, H. P. & Craven, B. M. (1988). *Energies of molecular-interactions from Bragg diffraction data*. *J. Am. Chem. Soc.* **110**, 775–782.
Stone, A. J. (1996). *The Theory of Intermolecular Forces*. New York: Oxford University Press.
Trickey, S. B., Alford, J. A. & Boettger, J. C. (2004). *Methods and implementation of robust, high-precision Gaussian basis DFT calculations for periodic systems: the GTOFF code*. In *Theoretical and Computational Chemistry*, edited by J. Leszcynski, Vol. 15, ch. 6. Amsterdam: Elsevier.
Volkov, A., Koritsansky, T. & Coppens, P. (2004). *Combination of the exact potential and multipole methods (ep/mm) for evaluation of intermolecular interaction energies with pseudoatom implementation of electron densities*. *Chem. Phys. Lett.* **391**, 170–175.
Wedemeyer, W. J., Arnautova, Y. A., Pillardy, J., Wawak, R. A., Czaplewsky, C. & Scheraga, H. A. (2000). *Reply to 'Comment on Crystal structure prediction by global optimization as a tool for evaluating potentials: role of the dipole moment correction terms in successful predictions'*. *J. Phys. Chem. B*, **104**, 8090–8092.
Williams, D. E. (2001). *Improved intermolecular force field for molecules containing H, C, N and O atoms, with application to nucleoside and peptide crystals*. *J. Comput. Chem.* **22**, 1154–1166.
Williams, D. E. & Cox, S. R. (1984). *Nonbonded potentials for azahydrocarbons: the importance of the Coulombic interactions*. *Acta Cryst.* B**40**, 404–417.
York, D. & Yang, W. (1994). *The fast Fourier Poisson (FFP) method for calculating Ewald sums*. *J. Chem. Phys.* **101**, 3298–3300.

International Tables for Crystallography (2010). Vol. B, Chapter 4.1, pp. 484–491.

4.1. Thermal diffuse scattering of X-rays and neutrons

By B. T. M. Willis

4.1.1. Introduction

Thermal motion of the atoms in a crystal gives rise to a reduction in the intensities of the Bragg reflections and to a diffuse distribution of non-Bragg scattering in the rest of reciprocal space. This distribution is known as thermal diffuse scattering (TDS). Measurement and analysis of TDS give information about the lattice dynamics of the crystal, *i.e.* about the small oscillatory displacements of the atoms from their equilibrium positions which arise from thermal excitations. Lattice-dynamical models form the basis for interpreting many physical properties – for example, specific heat and thermal conductivity – which cannot be explained by a static model of the crystal.

Reference to a lattice-dynamical model is found in Newton's *Principia*, which contains a discussion of the vibrations of a linear chain of equidistant mass points connected by springs. The model was used to estimate the speed of sound in air. The vibrational properties of a one-dimensional crystal treated as a linear chain of atoms provide the starting point for several modern treatises on the lattice dynamics of crystals.

The classical theory of the dynamics of three-dimensional crystals is based on the treatment of Born & von Kármán (1912, 1913). In this theory, the restoring force on an atom is determined not by the displacement of the atom from its equilibrium position, but by its displacement relative to its neighbours. The atomic motion is then considered in terms of travelling waves, or 'lattice vibrations', extending throughout the whole crystal. These waves are the normal modes of vibration, in which each mode is characterized by a wavevector $\mathbf{q}$, an angular frequency $\omega(\mathbf{q})$ and certain polarization properties.

For twenty years after its publication the Born–von Kármán treatment was eclipsed by the theory of Debye (1912). In the Debye theory the crystal is treated as a continuous medium instead of a discrete array of atoms. The theory gives a reasonable fit to the integral vibrational properties (for example, the specific heat or the atomic temperature factor) of simple monatomic crystals. It fails to account for the form of the frequency distribution function which relates the number of modes and their frequency.

An even simpler model than Debye's is due to Einstein (1907), who considered the atoms in the crystal to be vibrating independently of each other and with the same frequency $\omega_{\rm E}$. By quantizing the energy of each atom in units of $\hbar\omega_{\rm E}$, Einstein showed that the specific heat falls to zero at $T = 0$ K and rises asymptotically to the Dulong and Petit value for T much larger than $\hbar\omega_{\rm E}/k_{\rm B}$. ($\hbar$ is Planck's constant divided by 2π and $k_{\rm B}$ is Boltzmann's constant.) His theory accounts satisfactorily for the breakdown of equipartition of energy at low temperatures, but it predicts a more rapid fall-off of specific heat with decreasing temperature than is observed.

Deficiencies in the Debye theory were noted by Blackman (1937), who showed that they are overcome satisfactorily using the more rigorous Born–von Kármán theory. Extensive X-ray studies of Laval (1939) on simple structures such as sylvine, aluminium and diamond showed that the detailed features of the TDS could only be explained in terms of the Born–von Kármán theory. The X-ray work on aluminium was developed further by Olmer (1948) and by Walker (1956) to derive the phonon dispersion relations (see Section 4.1.5) along various symmetry directions in the crystal.

It is possible to measure the vibrational frequencies directly with X-rays, but such measurements are very difficult as lattice vibrational energies are many orders of magnitude less than X-ray energies. The situation is much more favourable with thermal neutrons because their wavelength is comparable with interatomic spacings and their energy is comparable with a quantum of vibrational energy (or phonon). The neutron beam is scattered inelastically by the lattice vibrations, exchanging energy with the phonons. By measuring the energy change for different directions of the scattered beam, the dispersion relations $\omega(\mathbf{q})$ can be determined. Brockhouse & Stewart (1958) reported the first dispersion curves to be derived in this way; since then the neutron technique has become the principal experimental method for obtaining detailed information about lattice vibrations.

In this chapter we shall describe briefly the standard treatment of the lattice dynamics of crystals. There follows a section on the theory of the scattering of X-rays by lattice vibrations, and a similar section on the scattering of thermal neutrons. We then refer briefly to experimental work with X-rays and neutrons. The final section is concerned with the measurement of elastic constants: these constants are required in calculating the TDS correction to measured Bragg intensities (see Section 7.4.2 of *IT* C, 2004).

4.1.2. Dynamics of three-dimensional crystals

For modes of vibration of very long wavelength, the crystal can be treated as a homogeneous elastic continuum without referring to its crystal or molecular structure. The theory of the propagation of these elastic waves is based on Hooke's law of force and on Newton's equations of motion. As the wavelength of the vibrations becomes shorter and shorter and approaches the separation of adjacent atoms, the calculation of the vibrational properties requires a knowledge of the crystal structure and of the nature of the forces between adjacent atoms. The three-dimensional treatment is based on the formulation of Born and von Kármán, which is discussed in detail in the book by Born & Huang (1954) and in more elementary terms in the books by Cochran (1973) and by Willis & Pryor (1975).

Before setting up the equations of motion, it is necessary to introduce three approximations:

(i) *The harmonic approximation*. When an atom is displaced from its equilibrium position, the restoring force is assumed to be proportional to the displacement, measured relative to the neighbouring atoms. The approximation implies no thermal expansion and other properties not possessed by real crystals; it is a reasonable assumption in the lattice-dynamical theory provided the displacements are not too large.

(ii) *The adiabatic approximation*. We wish to set up a potential function for the crystal describing the binding between the atoms. However, the binding involves electronic motions whereas the dynamics involve nuclear motions. The adiabatic approximation, known as the Born–Oppenheimer approximation in the context of molecular vibrations, provides the justification for adopting the same potential function to describe both the binding and the dynamics. Its essence is that the electronic and nuclear motions may be considered separately. This is possible if the nuclei move very slowly compared with the electrons: the electrons can then instantaneously take up a configuration appropriate to that of the displaced nuclei without changing their quantum state. The approximation holds well for insulators, where electronic transition energies are high owing to the large energy gap between filled and unfilled electron states. Surprisingly, it even works for

Copyright © 2010 International Union of Crystallography

metals, because (on account of the Pauli principle) only a few electrons near the Fermi level can make transitions.

(iii) *Periodic boundary conditions*. These are introduced to avoid problems associated with the free surface. The system is treated as an infinite crystal made up of contiguous, repeating blocks of the actual crystal. The periodic (or cyclic) boundary conditions require that the displacements of corresponding atoms in different blocks are identical. The validity of the conditions was challenged by Raman (1941*a*,*b*), but these objections were safely disposed of by Ledermann (1944).

4.1.2.1. Equations of motion

As a result of thermal fluctuations, the atoms vibrate about their equilibrium positions, so that the actual position of the κth atom in the lth primitive cell is given by

$$\mathbf{R}(\kappa l) = \mathbf{r}(\kappa l) + \mathbf{u}(\kappa l)$$

with $\mathbf{r}$ representing the equilibrium position and $\mathbf{u}$ the thermal displacement. (In lattice-dynamical theory it is advantageous to deal with the primitive cell, as it possesses the fewest degrees of freedom.) The kinetic energy of the vibrating crystal is

$$(1/2)\sum_{\kappa l\alpha} m(\kappa)\dot{u}_\alpha^2(\kappa l),$$

where $m(\kappa)$ is the mass of atom κ and the index α (α = 1, 2, 3) refers to the Cartesian components of the displacement. (The dot denotes the time derivative.)

If the adiabatic approximation is invoked, the potential energy V of the crystal can be expressed as a function of the instantaneous atomic positions. Expanding V in powers of $\mathbf{u}(\kappa l)$, using the three-dimensional form of Taylor's series, we have

$$V = V^{(0)} + V^{(1)} + V^{(2)} + V^{(3)} + \ldots,$$

where $V^{(0)}$ is the static (equilibrium) potential and $V^{(1)}$, $V^{(2)}$ are given by

$$V^{(1)} = \sum_{\kappa l\alpha} \left.\frac{\partial V}{\partial u_\alpha(\kappa l)}\right|_0 u_\alpha(\kappa l)$$

$$V^{(2)} = \tfrac{1}{2}\sum_{\kappa l\alpha}\sum_{\kappa' l'\alpha'} \left.\frac{\partial^2 V}{\partial u_\alpha(\kappa l)\partial u_{\alpha'}(\kappa' l')}\right|_0 u_\alpha(\kappa l)u_{\alpha'}(\kappa' l').$$

The subscript zero indicates that the derivatives are to be evaluated at the equilibrium configuration. In the harmonic approximation, $V^{(3)}$ and all higher terms in the expansion are neglected.

At equilibrium the forces on an atom must vanish, so that

$$V^{(1)} = 0.$$

Ignoring the static potential $V^{(0)}$, the quadratic term $V^{(2)}$ only remains and the Hamiltonian for the crystal (the sum of the kinetic and potential energies) is then

$$H = \tfrac{1}{2}\sum_{\kappa l\alpha} m(\kappa)\dot{u}_\alpha^2(\kappa l) + \tfrac{1}{2}\sum_{\kappa l\alpha}\sum_{\kappa' l'\alpha'} \Phi_{\alpha\alpha'}\begin{pmatrix}\kappa & \kappa'\\ l & l'\end{pmatrix} u_\alpha(\kappa l)u_{\alpha'}(\kappa' l'), \qquad (4.1.2.1)$$

where $\Phi_{\alpha\alpha'}$ is an element of the 3 × 3 'atomic force-constant matrix' and is defined (for distinct atoms κl, $\kappa' l'$) by

$$\Phi_{\alpha\alpha'}\begin{pmatrix}\kappa & \kappa'\\ l & l'\end{pmatrix} = \left.\frac{\partial^2 V}{\partial u_\alpha(\kappa l)\partial u_{\alpha'}(\kappa' l')}\right|_0.$$

It is the negative of the force in the α direction imposed on the atom (κl) when atom $(\kappa' l')$ is displaced unit distance along α' with all the remaining atoms fixed at their equilibrium sites. $\Phi_{\alpha\alpha'}$ is defined differently for the self-term with $\kappa = \kappa'$ and $l = l'$:

$$\Phi_{\alpha\alpha'}\begin{pmatrix}\kappa & \kappa\\ l & l\end{pmatrix} = -\sum_{\kappa l}\sum_{\substack{\kappa' l'\\ \kappa l\neq\kappa' l'}} \Phi_{\alpha\alpha'}\begin{pmatrix}\kappa & \kappa'\\ l & l'\end{pmatrix}.$$

Thus the self-matrix describes the force on (κl) when the atom itself is displaced with all the remaining atoms kept stationary.

There are restrictions on the number of distinct force constants $\Phi_{\alpha\alpha'}$: these are imposed by symmetry and by the requirement that the potential energy is invariant under infinitesimal translations and rotations of the rigid crystal. Such constraints are discussed in the book by Venkataraman *et al.* (1975).

Applying Hamilton's equations of motion to equation (4.1.2.1) now gives

$$m(\kappa)\ddot{u}_\alpha(\kappa l) = -\sum_{\kappa' l'\alpha'} \Phi_{\alpha\alpha'}\begin{pmatrix}\kappa & \kappa'\\ l & l'\end{pmatrix} u_{\alpha'}(\kappa' l'). \qquad (4.1.2.2)$$

These represent $3nN$ coupled differential equations, where n is the number of atoms per primitive cell ($\kappa = 1, \ldots, n$) and N is the number of cells per crystal ($l = 1, \ldots, N$).

By applying the periodic boundary conditions, the solutions of equation (4.1.2.2) can be expressed as running, or travelling, plane waves extending throughout the entire crystal. The number of independent waves (or normal modes) is $3nN$. Effectively, we have transferred to a new coordinate system: instead of specifying the motion of the individual atoms, we describe the thermal motion in terms of normal modes, each of which contributes to the displacement of each atom. The general solution for the α component of the displacement of (κl) is then given by the superposition of the displacements from all modes:

$$u_\alpha(\kappa l) = [m(\kappa)]^{-1/2}\sum_{j\mathbf{q}} |A_j(\mathbf{q})| e_\alpha(\kappa|j\mathbf{q}) \times \exp\{i[\mathbf{q}\cdot\mathbf{r}(\kappa l) - \omega_j(\mathbf{q})t]\}. \qquad (4.1.2.3)$$

Here $\mathbf{q}$ is the wavevector of a mode (specifying both its wavelength and direction of propagation in the crystal) and $\omega(\mathbf{q})$ its frequency. There are N distinct wavevectors, occupying a uniformly distributed mesh of N points in the Brillouin zone (reciprocal cell); each wavevector is shared by $3n$ modes which possess, in general, different frequencies and polarization properties. Thus an individual mode is conveniently labelled $(j\mathbf{q})$, where j is an index ($j = 1, \ldots, 3n$) indicating the *branch*. The scalar quantity $|A_j(\mathbf{q})|$ in equation (4.1.2.3) is the amplitude of excitation of $(j\mathbf{q})$ and $e_\alpha(\kappa|j\mathbf{q})$ is the element of the eigenvector $\mathbf{e}(j\mathbf{q})$ referring to the displacement in the α direction of the atom κ. The eigenvector itself, with dimensions $n \times 1$, determines the pattern of atomic displacements in the mode $(j\mathbf{q})$ and its magnitude is fixed by the orthonormality and closure conditions

$$\sum_{\alpha\kappa} e_\alpha^*(\kappa|j\mathbf{q})e_\alpha(\kappa|j'\mathbf{q}) = \delta_{jj'}$$

and

$$\sum_j e_\alpha(\kappa|j\mathbf{q})e^*_{\alpha'}(\kappa'|j\mathbf{q}) = \delta_{\alpha\alpha'}\delta_{\kappa\kappa'}$$

with * indicating complex conjugate and δ the Kronecker delta.

The pre-exponential, or amplitude, terms in (4.1.2.3) are independent of the cell number. This follows from Bloch's (1928) theorem which states that, for corresponding atoms in different cells, the motions are identical as regards their amplitude and direction and differ only in phase. The theorem introduces an enormous simplification as it allows us to restrict attention to the $3n$ equations of motion of the n atoms in just one cell, rather than the $3nN$ equations of motion for all the atoms in the crystal.

Substitution of (4.1.2.3) into (4.1.2.2) gives the equations of motion in the form

$$\omega_j^2(\mathbf{q})e_\alpha(\kappa|j\mathbf{q}) = \sum_{\alpha'\kappa'} D_{\alpha\alpha'}(\kappa\kappa'|\mathbf{q})e_{\alpha'}(\kappa'|j\mathbf{q}), \qquad (4.1.2.4)$$

in which $D_{\alpha\alpha'}$ is an element of the dynamical matrix $\mathbf{D}(\mathbf{q})$. $D_{\alpha\alpha'}$ is defined by

$$D_{\alpha\alpha'}(\kappa\kappa'|\mathbf{q}) = [m(\kappa)m(\kappa')]^{-1/2}\exp\{i\mathbf{q}[r(\kappa') - r(\kappa)]\} \times \sum \Phi_{\alpha\alpha'}\begin{pmatrix}\kappa & \kappa'\\ 0 & L\end{pmatrix}\exp[i\mathbf{q}\cdot\mathbf{r}(L)], \qquad (4.1.2.5)$$

where $\mathbf{r}(\kappa)$ is the position of atom κ with respect to the cell origin, L is $l' - l$ and $\mathbf{r}(L)$ is the separation between cells l and l'. The element $D_{\alpha\alpha'}$ is obtained by writing down the $\alpha\alpha'$ component of the force constant between atoms κ, κ' which are L cells apart and multiplying by the phase factor $\exp[i\mathbf{q}\cdot\mathbf{r}(L)]$; this term is then summed over those values of L covering the range of interaction of κ and κ'.

The dynamical matrix is Hermitian and has dimensions $3n \times 3n$. Its eigenvalues are the squared frequencies $\omega_j^2(\mathbf{q})$ of the normal modes and its eigenvectors $\mathbf{e}(j\mathbf{q})$ determine the corresponding pattern of atomic displacements. The frequencies of the modes in three of the branches, j, go to zero as $\mathbf{q}$ approaches zero: these are the acoustic modes. The remaining $3n - 3$ branches contain the optic modes. There are N distinct $\mathbf{q}$ vectors, and so, in all, there are $3N$ acoustic modes and $(3n - 3)N$ optic modes. Thus copper has acoustic modes but no optic modes, silicon and rock salt have an equal number of both, and lysozyme possesses predominantly optic modes.

4.1.2.2. Quantization of normal modes. Phonons

Quantum concepts are not required in solving the equations of motion (4.1.2.4) to determine the frequencies and displacement patterns of the normal modes. The only place where quantum mechanics is necessary is in calculating the energy of the mode, and from this the amplitude of vibration $|A_j(\mathbf{q})|$.

It is possible to discuss the theory of lattice dynamics from the beginning in the language of quantum mechanics (Donovan & Angress, 1971). Instead of treating the modes as running waves, they are conceived as an assemblage of indistinguishable quasi-particles called phonons. Phonons obey Bose–Einstein statistics and are not limited in number. The number of phonons, each with energy $\hbar\omega_j(\mathbf{q})$ in the vibrational state specified by $\mathbf{q}$ and j, is given by

$$n_j(\mathbf{q}) = \{\exp[\hbar\omega_j(\mathbf{q})/k_B T] - 1\}^{-1} \qquad (4.1.2.6)$$

and the mode energy $E_j(\mathbf{q})$ by

$$E_j(\mathbf{q}) = \hbar\omega_j(\mathbf{q})[n_j(\mathbf{q}) + (1/2)]. \qquad (4.1.2.7)$$

Thus the quantum number $n_j(\mathbf{q})$ describes the degree of excitation of the mode $(j\mathbf{q})$. The relation between $E_j(\mathbf{q})$ and the amplitude $|A_j(\mathbf{q})|$ is

$$E_j(\mathbf{q}) = N\omega_j^2(\mathbf{q})|A_j(\mathbf{q})|^2. \qquad (4.1.2.8)$$

Equations (4.1.2.6) to (4.1.2.8) together determine the value of $|A_j(\mathbf{q})|$ to be substituted into equation (4.1.2.3) to give the atomic displacement in terms of the absolute temperature and the properties of the normal modes.

In solving the lattice-dynamical problem using the Born–von Kármán analysis, the first step is to set up a force-constant matrix describing the interactions between all pairs of atoms. This is followed by the assembly of the dynamical matrix $\mathbf{D}$, whose eigenvalues give the frequencies of the normal modes and whose eigenvectors determine the patterns of atomic displacement for each mode.

Before considering the extension of this treatment to molecular crystals, we shall comment briefly on the less rigorous treatments of Einstein and Debye.

4.1.2.3. Einstein and Debye models

In the Einstein model it is assumed that each atom vibrates in its private potential well, entirely unaffected by the motion of its neighbours. There is no correlation between the motion of different atoms, whereas correlated motion – in the form of collective modes propagating throughout the crystal – is a central feature in explaining the characteristics of the TDS. Nevertheless, the Einstein model is occasionally used to represent modes belonging to flat optic branches of the dispersion relations, with the frequency written symbolically as $\omega(\mathbf{q}) = \omega_E$ (constant).

In the Debye model the optic branches are ignored. The dispersion relations for the remaining three acoustic branches are assumed to be the same and represented by

$$\omega(\mathbf{q}) = \mathbf{v}_s q, \qquad (4.1.2.9)$$

where $\mathbf{v}_s$ is a mean sound velocity. The Brillouin zone is replaced by a sphere with radius q_D chosen to ensure the correct number of modes. The linear relationship (4.1.2.9) holds right up to the boundary of the spherical zone. In an improved version of the Debye model, (4.1.2.9) is replaced by the expression

$$\omega(\mathbf{q}) = \mathbf{v}_s(2q_D/\pi)\sin(\pi q/2q_D), \qquad (4.1.2.10)$$

which is the same as (4.1.2.9) at $q = 0$ but gives a sinusoidal dispersion relation with zero slope at the zone boundary.

4.1.2.4. Molecular crystals

The full Born–von Kármán treatment becomes excessively cumbersome when applied to most molecular crystals. For example, for naphthalene with two molecules or 36 atoms in the primitive cell, the dynamical matrix has dimensions 108×108. Moreover, the physical picture of molecules or of groups of atoms, vibrating in certain modes as quasi-rigid units, is lost in the full treatment.

To simplify the setting up of the dynamical matrix, it is assumed that the molecules vibrate as rigid units in the crystal with each molecule possessing three translational and three rotational (librational) degrees of freedom. The motion of these rigid groups as a whole is described by the *external modes* of motion, whereas the *internal modes* arise from distortions within

an individual group. The frequencies of these internal modes, which are largely determined by the strong intramolecular forces, are unaffected by the phase of the oscillation between neighbouring cells: the modes are taken, therefore, to be equivalent to those of the free molecule. The remaining external modes are calculated by applying the Born–von Kármán procedure to the crystal treated as an assembly of rigid molecules.

The dynamical matrix $\mathbf{D}(\mathbf{q})$ now has dimensions $6n' \times 6n'$, where n' is the number of molecules in the primitive cell: for naphthalene, $\mathbf{D}$ is reduced to 12×12. The elements of $\mathbf{D}$ can be expressed in the same form as equation (4.1.2.5) for an atomic system. κ, κ' refer to molecules which are L cells apart and the indices α, α' $(= 1, \ldots, 6)$ label the six components of translation and rotation. $m(\kappa)$ in equation (4.1.2.5) is replaced by $m_\alpha(\kappa)$ where m_α represents the 3×3 molecular-mass matrix for $\alpha = 1, 2, 3$ and represents the 3×3 moment-of-inertia matrix referred to the principal axes of inertia for $\alpha = 4, 5, 6$. The 6×6 force-torque constant matrices $\Phi_{\alpha\alpha'}$ are derived by taking the second derivative of the potential energy of the crystal with respect to the coordinates of translation and rotation.

4.1.3. Scattering of X-rays by thermal vibrations

The change of frequency, or energy, of X-rays on being scattered by thermal waves is extremely small. The differential scattering cross section, $\mathrm{d}\sigma/\mathrm{d}\Omega$, giving the probability that X-rays are scattered into the solid angle $\mathrm{d}\Omega$ is then

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \langle |Y(Q)|^2 \rangle,$$

where

$$Y(Q) = \sum_{\kappa l} f_\kappa(\mathbf{Q}) \exp\{i\mathbf{Q} \cdot [\mathbf{r}(\kappa l) + \mathbf{u}(\kappa l)]\}. \qquad (4.1.3.1)$$

The angled brackets indicate an average value over a period of time much longer than the period of oscillation of an atom. $\mathbf{Q}$ is the 'scattering vector' defined by

$$\mathbf{Q} = \mathbf{k} - \mathbf{k}_0,$$

where $\mathbf{k}$ and $\mathbf{k}_0$ are the wavevectors (each of magnitude $2\pi/\lambda$) of the scattered and incident beams, respectively. For Bragg scattering the magnitude of $\mathbf{Q}$ is $4\pi(\sin\theta)/\lambda$, where θ is the Bragg angle. $f_\kappa(\mathbf{Q})$ is the scattering factor of the κth atom in the unit cell.

The cross section $\mathrm{d}\sigma/\mathrm{d}\Omega$ can be expanded as a power series:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(0)} + \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(1)} + \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(2)} + \ldots. \qquad (4.1.3.2)$$

The individual terms on the right-hand side refer to the cross sections for zero-order, first-order, second-order scattering ..., *i.e.* for processes involving no exchange of energy between the incident radiation and the crystal (Bragg scattering), the exchange of one quantum of lattice vibrational energy (one-phonon scattering), the exchange of two quanta (two-phonon scattering) ....

The instantaneous thermal displacement $\mathbf{u}(\kappa l)$ of the atom (κl) can be expressed, using equations (4.1.2.3), (4.1.2.6), (4.1.2.7) and (4.1.2.8), as a superposition of the displacements of the $3nN$ ($\sim 10^{23}$) independent normal modes of vibration:

$$\mathbf{u}(\kappa l) = [Nm(\kappa)]^{-1/2} \sum_{j\mathbf{q}} \frac{E_j^{-1/2}(\mathbf{q})}{\omega_j^2(\mathbf{q})} \mathbf{e}(\kappa|j\mathbf{q}) \times \exp\{i[\mathbf{q} \cdot \mathbf{r}(\kappa l) - \omega_j(\mathbf{q})t]\}. \qquad (4.1.3.3)$$

Explicit expressions can now be given for the partial cross sections in equation (4.1.3.2). The cross section for Bragg scattering is

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(0)} = N|F(\mathbf{Q})|^2 \frac{(2\pi)^3}{\mathbf{v}} \sum_{\mathbf{h}} \delta(\mathbf{Q} - \mathbf{Q}_\mathbf{h}), \qquad (4.1.3.4)$$

where $\mathbf{v}$ is the cell volume. The delta function requires that

$$\mathbf{Q} = \mathbf{Q}_\mathbf{h},$$

where $\mathbf{Q}_\mathbf{h}$ is a reciprocal-lattice vector, so that scattering is restricted to the points $\mathbf{h}$ of the reciprocal lattice. The structure factor $F(\mathbf{Q})$ is

$$F(\mathbf{Q}) = \sum_\kappa f_\kappa(\mathbf{Q}) \exp[i\mathbf{Q} \cdot \mathbf{r}(\kappa)] \exp(-W_\kappa),$$

where the exponent W_κ of the temperature factor of the atom κ is calculated by summing over the normal modes:

$$W_\kappa = \frac{1}{2Nm(\kappa)} \sum_{j\mathbf{q}} |\mathbf{Q}_h \cdot \mathbf{e}(\kappa|j\mathbf{q})|^2 \frac{E_j(\mathbf{q})}{\omega_j^2(\mathbf{q})}. \qquad (4.1.3.5)$$

The last equation shows that the acoustic modes, with frequencies approaching zero as $\mathbf{q} \to 0$, make the largest contribution to the temperature factor.

The one-phonon cross section is

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(1)} = \frac{(2\pi)^3}{\mathbf{v}} \sum_{j\mathbf{q}} \frac{|G_j(\mathbf{Q}, \mathbf{q})|^2 E_j(\mathbf{q})}{\omega_j^2(\mathbf{q})} \times \sum_\mathbf{h} \delta(\mathbf{Q} \pm \mathbf{q} - \mathbf{Q}_\mathbf{h}), \qquad (4.1.3.6)$$

where $G_j(\mathbf{Q}, \mathbf{q})$ is the 'structure factor for one-phonon scattering' by the mode $(j\mathbf{q})$ and is given by

$$G_j(\mathbf{Q}, \mathbf{q}) = \sum_\kappa \frac{\mathbf{Q} \cdot \mathbf{e}(\kappa|j\mathbf{q})}{[m(\kappa)]^{1/2}} \cdot f_\kappa(\mathbf{Q}) \times \exp[i\mathbf{Q} \cdot \mathbf{r}(\kappa)] \exp(-W_\kappa).$$

The delta function in equation (4.1.3.6) implies that

$$\mathbf{Q} = \mathbf{Q}_\mathbf{h} \pm \mathbf{q}, \qquad (4.1.3.7)$$

so that the scattering from the $3n$ modes with the same wavevector $\mathbf{q}$ is restricted to pairs of points in reciprocal space which are displaced by $\pm\mathbf{q}$ from the reciprocal-lattice points. (These satellite reflections are analogous to the pairs of 'ghosts' near the principal diffraction maxima in a grating ruled with a periodic error.) There is a huge number, N, of $\mathbf{q}$ vectors which are uniformly distributed throughout the Brillouin zone, and each of these vectors gives a cross section in accordance with equation (4.1.3.6). Thus the one-phonon TDS is spread throughout the

whole of reciprocal space, rising to a maximum at the reciprocal-lattice points where $\omega \rightarrow 0$ for the acoustic modes.

For two-phonon scattering, involving modes with wavevectors $\mathbf{q}_1$ and $\mathbf{q}_2$, the scattering condition becomes

$$\mathbf{Q} = \mathbf{Q}_\mathbf{h} \pm \mathbf{q}_1 \pm \mathbf{q}_2. \qquad (4.1.3.8)$$

The intensity at any point in reciprocal space is now contributed by a very large number of pairs of elastic waves with wavevectors satisfying equation (4.1.3.8). These vectors span the entire Brillouin zone, and so the variation of the two-phonon intensity with location in reciprocal space is less pronounced than for one-phonon scattering.

Expressions for $(\mathrm{d}\sigma/\mathrm{d}\Omega)^{(2)}$ and for higher terms in equation (4.1.3.2) will not be given, but a rough estimate of their relative magnitudes can be derived by using the Einstein model of the crystal. All frequencies are the same, $\omega_j(\mathbf{q}) = \omega_\mathrm{E}$, and for one atom per cell ($n = 1$) the exponent of the temperature factor, equation (4.1.3.5), becomes

$$W = \frac{Q^2 k_\mathrm{B} T}{2m\omega_\mathrm{E}^2}, \qquad (4.1.3.9)$$

assuming classical equipartition of energy between modes: $E_j(\mathbf{q}) = k_\mathrm{B}T$. The cross sections for zero-order, first-order, second-order ... scattering are then

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(0)} = Nf^2 \exp(-2W)$$
$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(1)} = Nf^2 \exp(-2W)2W$$
$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)^{(2)} = Nf^2 \exp(-2W)\tfrac{1}{2}(2W)^2$$
$$\vdots$$

and the total cross section is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = Nf^2 \exp(-2W)\big\{1 + 2W + \tfrac{1}{2}(2W)^2 + \tfrac{1}{6}(2W)^3 + \ldots\big\}.$$

The expression in curly brackets is the expansion of $\exp(2W)$. The nth term in the expansion, associated with the nth-order (n-phonon) process, is proportional to W^n or to $Q^{2n}T^n$. The higher-order processes are more important, therefore, at higher values of $(\sin\theta)/\lambda$ and at higher temperatures.

Our treatment so far applies to the TDS from single crystals. It can be extended to cover the TDS from polycrystalline samples, but the calculations are more complicated as the first-order scattering at a fixed value of $(\sin\theta)/\lambda$ is contributed by phonon wavevectors extending over the whole of the Brillouin zone. For a fuller discussion of the TDS from powders see Section 7.4.2 in *IT* C (2004).

4.1.4. Scattering of neutrons by thermal vibrations

The amplitude of the X-ray beam scattered by a single atom is denoted by the atomic scattering factor $f(\mathbf{Q})$. The corresponding quantity in neutron scattering is the scattering amplitude or scattering length b of an atom. b is independent of scattering angle and is also independent of neutron wavelength apart from for a few isotopes (*e.g.* ^{113}Cd, ^{149}Sm) with resonances in the thermal neutron region.

The scattering amplitude for atoms of the same chemical element can vary randomly from one atom to the next, as different amplitudes are associated with different isotopes. If the nucleus has a nonzero spin, even a single isotope has two different amplitudes, dependent on whether the nuclear spin is parallel or antiparallel to the spin of the incident neutron. If there is a variation in the amplitude associated with a particular type of atom, some of the waves scattered by the atoms will interfere with one another and some will not. The first part is called coherent scattering and the second incoherent scattering. The amplitude of the coherent scattering is determined by the mean atomic scattering amplitude, averaged over the various isotopes and spin states of the atom, and is known as the coherent scattering length, $\bar{b}$.

A crucial difference between neutrons and X-rays concerns their energies:

$$E = \hbar^2 k^2/2m_\mathrm{n} \quad \text{(neutrons)}$$
$$= c\hbar k \quad \text{(X-rays)},$$

where m_n is the neutron mass and c the velocity of light. At $\lambda = 1$ Å, neutrons have an energy of 0.08 eV or a temperature of about 800 K; for X-rays of this wavelength the corresponding temperature exceeds 10^8 K! Thermal neutrons have energies comparable with phonon energies, and so inelastic scattering processes, involving the exchange of energy between neutrons and phonons, produce appreciable changes in neutron energy. (These changes are readily determined from the change in wavelength or velocity of the scattered neutrons.) It is customary to refer to the thermal diffuse scattering of neutrons as 'inelastic neutron scattering' to draw attention to this energy change.

For a scattering process in which energy is exchanged with just one phonon, energy conservation gives

$$E_0 - E = \varepsilon\hbar\omega_j(\mathbf{q}), \qquad (4.1.4.1)$$

where E_0 and E are the energies of the neutron before and after scattering and $\varepsilon = \pm 1$. If $\varepsilon = +1$ the neutron loses energy by creating a phonon ('phonon emission'), and if $\varepsilon = -1$ it gains energy by annihilating a phonon ('phonon absorption').

The partial differential scattering cross section $\mathrm{d}^2\sigma/\mathrm{d}\Omega\,\mathrm{d}\omega$ gives the probability that neutrons will be scattered into a small solid angle $\mathrm{d}\Omega$ about the direction $\mathbf{k}$ with a change of energy between $\hbar\omega$ and $\hbar(\omega + \mathrm{d}\omega)$. This cross section can be split into two terms, known as the coherent and incoherent cross sections:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega} = \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)_\mathrm{coh} + \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)_\mathrm{incoh}.$$

The coherent cross section depends on the correlation between the positions of all the atoms at different times, and so gives interference effects. The incoherent cross section depends only on the correlation between the positions of the same atom at different times, giving no interference effects. Incoherent inelastic scattering is the basis of a powerful technique for studying the dynamics of molecular crystals containing hydrogen (Boutin & Yip, 1968).

The coherent scattering cross section $(\mathrm{d}^2\sigma/\mathrm{d}\Omega\,\mathrm{d}\omega)_\mathrm{coh}$ can be expanded, as in the X-ray case [equation (4.1.3.2)], into terms representing the contributions from zero-phonon, one-phonon, two-phonon ... scattering. To determine phonon dispersion relations, we measure the one-phonon contribution and this arises from both phonon emission and phonon absorption:

$$\left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)^{(1)}_{\mathrm{coh}} = \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)^{(1)}_{\mathrm{coh}+1} + \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)^{(1)}_{\mathrm{coh}-1}.$$

The superscript (1) denotes a one-phonon process, and the subscript +1 (−1) indicates emission (absorption).

The emission cross section is given by (Squires, 1978)

$$\begin{aligned}\left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)^{(1)}_{\mathrm{coh}+1} = {} & \frac{\hbar k}{k_0}\frac{4\pi^3}{\mathbf{v}}\sum_{j\mathbf{q}}\sum_{\mathbf{Q_h}}\frac{[n_j(\mathbf{q})+1]}{\omega_j(\mathbf{q})} \\ & \times \left|\frac{b_\kappa}{[m(\kappa)]^{1/2}}\exp[i\mathbf{Q}\cdot\mathbf{r}(\kappa)]\right. \\ & \left.\times\,\mathbf{Q}\cdot\mathbf{e}(\kappa|j\mathbf{q})\exp(-W_\kappa)\right|^2 \\ & \times\,\delta[\omega-\omega_j(\mathbf{q})]\delta(\mathbf{Q}-\mathbf{Q_h}+\mathbf{q}),\end{aligned} \tag{4.1.4.2}$$

whereas for phonon absorption

$$\begin{aligned}\left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)^{(1)}_{\mathrm{coh}-1} = {} & \frac{\hbar k}{k_0}\frac{4\pi^3}{\mathbf{v}}\sum_{j\mathbf{q}}\sum_{\mathbf{Q_h}}\frac{[n_j(\mathbf{q})]}{\omega_j(\mathbf{q})} \\ & \times \left|\frac{b_\kappa}{[m(\kappa)]^{1/2}}\exp[i\mathbf{Q}\cdot\mathbf{r}(\kappa)]\right. \\ & \left.\times\,\mathbf{Q}\cdot\mathbf{e}(\kappa|j\mathbf{q})\exp(-W_\kappa)\right|^2 \\ & \times\,\delta[\omega+\omega_j(\mathbf{q})]\delta(\mathbf{Q}-\mathbf{Q_h}-\mathbf{q}).\end{aligned} \tag{4.1.4.3}$$

The first delta function in these two expressions embodies energy conservation,

$$\hbar\omega \equiv \frac{\hbar^2}{2m_{\mathrm{n}}}(k_0^2-k^2) = \varepsilon\hbar\omega_j(\mathbf{q}),$$

and the second embodies conservation of momentum,

$$\mathbf{Q} \equiv \mathbf{k}-\mathbf{k}_0 = \mathbf{Q_h}-\varepsilon\mathbf{q}. \tag{4.1.4.4}$$

The phonon population number, $n_j(\mathbf{q})$, tends to zero as $T \to 0$ [see equation (4.1.2.6)], so that the one-phonon absorption cross section is very small at low temperatures where there are few phonons for the radiation to absorb. Comparison of equations (4.1.4.2) and (4.1.4.3) shows that there is always a greater probability that the neutrons are scattered with energy loss rather than with energy gain.

Normally it is not possible in X-ray experiments to distinguish between phonon emission and phonon absorption, and the measured cross section is obtained by summing over all energy transfers. The cross section for X-rays can be derived from the neutron formulae, equations (4.1.4.2) and (4.1.4.3), by putting $k = k_0$ and by replacing b_κ with $f_\kappa(\mathbf{Q})$. Integration over ω and addition of the parts for emission and absorption gives the X-ray formula (4.1.3.6).

The theory of neutron scattering can also be formulated in terms of thermal averages known as Van Hove correlation functions (Van Hove, 1954). For example, the partial differential cross section for coherent scattering is

$$\left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}\omega}\right)_{\mathrm{coh}} = \frac{k}{k_0}S(\mathbf{Q},\omega),$$

where

$$S(\mathbf{Q},\omega) = \frac{1}{2\pi\hbar}\int G(\mathbf{r},t)\exp[i(\mathbf{Q}\cdot\mathbf{r}-\omega t)]\,\mathrm{d}\mathbf{r}\,\mathrm{d}t.$$

$S(\mathbf{Q},\omega)$ is the Fourier transform in space and time of $G(\mathbf{r},t)$, the time-dependent pair-correlation function. The classical interpretation of $G(\mathbf{r},t)$ is that it is the probability of finding any atom at time t in a volume $\mathrm{d}\mathbf{r} = \mathrm{d}^3r$, if there is an atom at the origin at time zero.

4.1.5. Phonon dispersion relations

Both X-rays and neutrons are used for determining crystal structures, but the X-ray method plays the dominant role. The reverse is true for the measurement of phonon dispersion relations: the experimental determination of $\omega(\mathbf{q})$ *versus* $\mathbf{q}$ was first undertaken with X-rays, but the method has been superseded by the technique of coherent inelastic neutron scattering (or neutron spectroscopy). For phonon wavevectors lying anywhere within the first Brillouin zone, it is necessary to employ radiation of wavelength comparable with interatomic distances and of energy comparable with lattice vibrational energies. X-rays satisfy the first of these conditions, but not the second, whereas the opposite holds for infrared radiation. Thermal neutrons satisfy both conditions simultaneously.

4.1.5.1. Measurement with X-rays

Frequencies can be derived indirectly with X-rays from the intensity of the thermal diffuse scattering. For a monatomic crystal with one atom per primitive cell, there are no optic modes and the one-phonon TDS intensity, equation (4.1.3.6), reduces to

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = Q^2f^2e^{-2W}\sum_{j=1}^{3}\frac{E_j(\mathbf{q})}{\omega^2(\mathbf{q})}\cos^2\alpha_j(\mathbf{q})\delta(\mathbf{Q}\pm\mathbf{q}-\mathbf{Q_h}), \tag{4.1.5.1}$$

where $\alpha_j(\mathbf{q})$ is the angle between $\mathbf{Q}$ and the direction of polarization of the mode $(j\mathbf{q})$. There are three acoustic modes associated with each wavevector $\mathbf{q}$, but along certain directions of $\mathbf{Q}$ it is possible to isolate the intensities contributed by the individual modes by choosing $\alpha_j(\mathbf{q})$ to be close to 0 or 90°. Equation (4.1.5.1) can then be employed to derive the frequency $\omega_j(\mathbf{q})$ for just one mode. The measured intensity must be corrected for multiphonon and Compton scattering, both of which can exceed the intensity of the one-phonon scattering. The correction for two-phonon scattering involves an integration over the entire Brillouin zone, and this in turn requires an approximate knowledge of the dispersion relations. The correction for Compton scattering can be made by repeating the measurements at low temperature.

This X-ray method is hardly feasible for systems with several atoms in the primitive cell. It comes into its own for those few materials which cannot be examined by neutrons. These include boron, cadmium and samarium with high absorption cross sections for thermal neutrons, and vanadium with a very small coherent (and a large incoherent) cross section for the scattering of neutrons. An important feature of TDS measurements with X-rays is in providing an independent check on interatomic or intermolecular force constants derived from measurements with inelastic neutron scattering. The force model is used to generate

phonon frequencies and eigenvectors, which are then employed to compute the one-phonon and multi-phonon contributions to the X-ray TDS. Any discrepancy between calculated and observed X-ray intensities might be ascribed to such features as ionic deformation (Buyers *et al.*, 1968) or anharmonicity (Schuster & Weymouth, 1971).

With the availability of dedicated beamlines at third-generation synchrotrons, inelastic X-ray scattering is an alternative method of determining phonon dispersion relations in crystals (Burkel, 2000). The basic design of the X-ray instrument is the same as the three-axis spectrometer for neutron scattering (see Section 4.1.5.2 below): however, it is necessary to measure the minute change in energy of the scattered X-rays to a resolution of a few meV. This *tour de force* is achieved by adopting some novel experimental procedures. These include the use of a monochromator and analyser of perfect silicon crystals, both of which are operated under extreme conditions of Bragg back reflection ($\theta \simeq 89.95^\circ$). Also, energy scans are performed by varying the relative temperature (and thereby the relative lattice spacing) between the monochromator and analyser with millikelvin precision. The X-ray spot size is in the submillimetre range, and so very small samples can be examined. A good example of this X-ray method is the study by Ruf *et al.* (2001) of the optic and acoustic modes of GaN, which possesses the wurtzite structure: crystals of GaN are too small to be examined by the traditional neutron method.

4.1.5.2. Measurement with neutrons

The inelastic scattering of neutrons by phonons gives rise to changes of energy which are readily measured and converted to frequencies $\omega_j(\mathbf{q})$ using equation (4.1.4.1). The corresponding wavevector $\mathbf{q}$ is derived from the momentum conservation relation (4.1.4.4). Nearly all phonon dispersion relations determined to date have been obtained in this way. Well over 200 materials have been examined, including half the chemical elements, a large number of alloys and diatomic compounds, and rather fewer molecular crystals (Dolling, 1974; Bilz & Kress, 1979). Phonon dispersion curves have been determined in crystals with up to ten atoms in the primitive cell, for example, tetracyanoethylene (Chaplot *et al.*, 1983).

The principal instrument for determining phonon dispersion relations with neutrons is the triple-axis spectrometer, first designed and built by Brockhouse (Brockhouse & Stewart, 1958). The modern instrument is unchanged apart from running continuously under computer control. A beam of thermal neutrons falls on a single-crystal monochromator, which Bragg reflects a single wavelength on to the sample in a known orientation. The magnitude of the scattered wavevector, and hence the change of energy on scattering by the sample, is found by measuring the Bragg angle at which the neutrons are reflected by the crystal analyser. The direction of $\mathbf{k}$ is defined by a collimator between the sample and analyser.

In the 'constant $\mathbf{Q}$' mode of operating the triple-axis spectrometer, the phonon wavevector is kept fixed while the energy transfer $\hbar\omega$ is varied. This allows the frequency spectrum to be determined for all phonons sharing the same $\mathbf{q}$; the spectrum will contain up to $3n$ frequencies, corresponding to the $3n$ branches of the dispersion relations.

In an inelastic neutron scattering experiment, where the TDS intensity is of the order of one-thousandth of the Bragg intensity, it is necessary to use a large sample with a volume of 1 cm^3, or more. The sample should have a high cross section for coherent scattering as compared with the cross sections for incoherent scattering and for true absorption. Crystals containing hydrogen should be deuterated.

Dolling (1974) has given a comprehensive review of the measurement of phonon dispersion relations by neutron spectroscopy.

4.1.5.3. Interpretation of dispersion relations

The usual procedure for analysing dispersion relations is to set up the Born–von Kármán formalism with an assumed set of interatomic force constants Φ. The calculated frequencies $\omega_j(\mathbf{q})$ are then derived from the eigenvalues of the dynamical matrix $\mathbf{D}$ (Section 4.1.2.1) and the force constants fitted, by least squares, to the observed frequencies. Several sets of force constants may describe the frequencies equally well, and to decide which set is preferable it is necessary to compare eigenvectors as well as eigenvalues (Cochran, 1971).

The main interest in the curves is in testing different models of interatomic potentials, whose derivatives are related to the interatomic force constants. For the solid inert gases the curves are reproduced reasonably well using a two-parameter Lennard–Jones 6–12 potential, although calculated frequencies are systematically higher than the experimental points near the Brillouin-zone boundary (Fujii *et al.*, 1974). To reproduce the dispersion relations in metals it is necessary to use a large number of interatomic force constants, extending to at least fifth neighbours. The number of independent constants is then too large for a meaningful analysis with the Born–von Kármán theory, but in the pseudo-potential approximation (Harrison, 1966) only two parameters are required to give good agreement between calculated and observed frequencies of simple metals such as aluminium. In the rigid-ion model for ionic crystals, the ions are treated as point charges centred on the nuclei and polarization of the outermost electrons is ignored. This is unsatisfactory at high frequencies. In the shell model, polarization is accounted for by representing the ion as a rigid core connected by a flexible spring to a polarizable shell of outermost electrons. There are many variants of this model – extended shell, overlap shell, deformation dipole, breathing shell ... (Bilz & Kress, 1979). For molecular crystals the contributions to the force constants from the intermolecular forces can be derived from the nonbonded atomic pair potential of, say, the 6-exponential type:

$$\varphi(r) = -\frac{A_{ij}}{r^6} + B_{ij}\exp(-C_{ij}r).$$

Here, i, j label atoms in different molecules. The values of the parameters A, B, C depend on the pair of atomic species i, j only. For hydrocarbons they have been tabulated for different atom pairs by Kitaigorodskii (1966) and Williams (1967). The 6-exponential potential is applicable to molecular crystals that are stabilized mainly by London–van der Waals interactions; it is likely to fail when hydrogen bonds are present.

4.1.6. Measurement of elastic constants

There is a close connection between the theory of lattice dynamics and the theory of elasticity. Acoustic modes of vibration of long wavelength propagate as elastic waves in a continuous medium, with all the atoms in one unit cell moving in phase with one another. These vibrations are sound waves with velocities which can be calculated from the macroscopic elastic constants and from the density. The sound velocities are also given by the slopes, at $q = 0$, of the acoustic branches of the phonon dispersion relations. A knowledge of the velocities, or of the elastic constants, is necessary in estimating the TDS contribution to measured Bragg intensities.

The elastic constants relate the nine components of stress and nine components of strain, making 81 constants in all. This large number is reduced to 36, because there are only six independent components of stress and six independent components of strain. An economy of notation is now possible, replacing the indices 11 by 1, 22 by 2, 33 by 3, 23 by 4, 31 by 5 and 12 by 6, so that the

elastic constants are represented by c_{ij} with $i, j = 1, \ldots, 6$. Applying the principle of conservation of energy gives

$$c_{ij} = c_{ji}$$

and the number of constants is reduced further to 21. Crystal symmetry effects yet a further reduction. For cubic crystals there are just three independent constants (c_{11}, c_{12}, c_{44}) and the 6×6 matrix of elastic constants is

$$\begin{pmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{pmatrix}.$$

In principle, the elastic constants can be derived from static measurements of the four quantities – compressibility, Poisson's ratio, Young's modulus and rigidity modulus. The measurements are made along different directions in the crystal: at least six directions are needed for the orthorhombic system. The accuracy of the static method is limited by the difficulty of measuring small strains.

Dynamic methods are more accurate as they depend on measuring a frequency or velocity. For a cubic crystal, the three elastic constants can be derived from the three sound velocities propagating along the single direction [110]; for non-cubic crystals the velocities must be measured along a number of non-equivalent directions.

Sound velocities can be determined in a number of ways. In the ultrasonic pulse technique, a quartz transducer sends a pulse through the crystal; the pulse is reflected from the rear surface back to the transducer, and the elapsed time for the round trip of several cm is measured. Brillouin scattering of laser light is also used (Vacher & Boyer, 1972). Fluctuations in dielectric constant caused by (thermally excited) sound waves give rise to a Doppler shift of the light frequency. The sound velocity is readily calculated from this shift, and the elastic constants are then obtained from the velocities along several directions, using the Christoffel relations (Hearmon, 1956). The Brillouin method is restricted to transparent materials. This restriction does not apply to neutron diffraction methods, which employ the inelastic scattering of neutrons (Willis, 1986; Schofield & Willis, 1987; Popa & Willis, 1994).

Tables of elastic constants of cubic and non-cubic crystals have been compiled by Hearmon (1946, 1956) and by Huntingdon (1958).

References

Bilz, H. & Kress, W. (1979). *Phonon Dispersion Relations in Insulators*. Berlin: Springer-Verlag.

Blackman, M. (1937). *On the vibrational spectrum of a three-dimensional lattice. Proc. R. Soc. London Ser. A*, **159**, 416–431.

Bloch, F. (1928). *Über die Quantenmechanik der Elektronen in Kristallgittern. Z. Phys.* **52**, 555–600.

Born, M. & Huang, K. (1954). *The Dynamical Theory of Crystal Lattices.* Oxford: Clarendon Press.

Born, M. & von Kármán, T. (1912). *Über Schwingungen in Raumgittern. Phys. Z.* **13**, 279–309.

Born, M. & von Kármán, T. (1913). *Zur Theorie der spezifischen Wärme. Phys. Z.* **14**, 15–19.

Boutin, H. & Yip, S. (1968). *Molecular Spectroscopy with Neutrons.* Cambridge: MIT Press.

Brockhouse, B. N. & Stewart, A. T. (1958). *Normal modes of aluminum by neutron spectrometry. Rev. Mod. Phys.* **30**, 236–249.

Burkel, E. (2000). *Phonon spectroscopy and inelastic X-ray scattering. Rep. Prog. Phys.* **63**, 171–232.

Buyers, W. J. L., Pirie, J. D. & Smith, T. (1968). *X-ray scattering from deformable ions. Phys. Rev.* **165**, 999–1005.

Chaplot, S. L., Mierzejewski, A., Pawley, G. S., Lefebvre, J. & Luty, T. (1983). *Phonon dispersion of the external and low-frequency internal vibrations in monoclinic tetracyanoethylene. J. Phys. C*, **16**, 625–644.

Cochran, W. (1971). *The relation between phonon frequencies and interatomic force constants. Acta Cryst.* A**27**, 556–559.

Cochran, W. (1973). *The Dynamics of Atoms in Crystals.* London: Edward Arnold.

Debye, P. (1912). *Zur Theorie der spezifischen Wärme. Ann. Phys. (Leipzig)*, **39**, 789–839.

Dolling, C. (1974). *Dynamical Properties of Solids*, Vol. 1, edited by G. K. Horton & A. A. Maradudin, pp. 541–629. Amsterdam: North-Holland.

Donovan, B. & Angress, J. F. (1971). *Lattice Vibrations.* London: Chapman and Hall.

Einstein, A. (1907). *Die Plancksche Theorie der Strahlung und die Theorie der spezifisichen Wärme. Ann. Phys. (Leipzig)*, **22**, 180–190, 800.

Fujii, Y., Lurie, N. A., Pynn, R. & Shirane, G. (1974). *Inelastic neutron scattering from solid ^{36}Ar. Phys. Rev. B*, **10**, 3647–3659.

Harrison, W. A. (1966). *Phonons in Perfect Lattices*, edited by R. W. H. Stevenson, pp. 73–109. Edinburgh: Oliver & Boyd.

Hearmon, R. F. S. (1946). *The elastic constants of anisotropic materials. Rev. Mod. Phys.* **18**, 409–440.

Hearmon, R. F. S. (1956). *The elastic constants of anisotropic materials. II. Adv. Phys.* **5**, 323–382.

Huntingdon, H. B. (1958). *The elastic constants of crystals. Solid State Phys.* **7**, 213–351.

International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

Kitaigorodskii, A. J. (1966). *Empilement des molécules dans un cristal, potentiel d'interaction des atomes non liés par des liaisons de valence, et calcul du mouvement des molécules. J. Chim. Phys.* **63**, 8–16.

Laval, J. (1939). *Etude éxperimentale de la diffusion des rayons X par les cristaux. Bull. Soc. Fr. Minéral.* **62**, 137–253.

Ledermann, W. (1944). *Asymptotic formulae relating to the physical theory of crystals. Proc. R. Soc. London Ser. A*, **82**, 362–377.

Olmer, P. (1948). *Dispersion des vitesses des ondes acoustiques dans l'aluminium. Acta Cryst.* **1**, 57–63.

Popa, N. C. & Willis, B. T. M. (1994). *Thermal diffuse scattering in time-of-flight neutron diffractometry. Acta Cryst.* A**50**, 57–63.

Raman, C. V. (1941*a*). *The quantum theory of X-ray reflection. Proc. Indian Acad. Sci. Sect. A,* **14**, 317–376.

Raman, C. V. (1941*b*). *The thermal energy of crystalline solids: basic theory. Proc. Indian Acad. Sci. Sect. A,* **14**, 459–467.

Ruf, T., Serrano, J., Cardona, M., Pavone, P., Pabst, M., Krisch, M., D'Astuto, M., Suski, T., Grzegory, I. & Leszczynski, M. (2001). *Phonon dispersion curve in wurtzite-structure GaN determined by inelastic x-ray scattering. Phys. Rev. Lett.* **86**, 906–909.

Schofield, P. & Willis, B. T. M. (1987). *Thermal diffuse scattering in time-of-flight neutron diffraction. Acta Cryst.* A**43**, 803–809.

Schuster, S. L. & Weymouth, J. W. (1971). *Study of thermal diffuse X-ray scattering from lead single crystals. Phys. Rev. B*, **3**, 4143–4153.

Squires, G. L. (1978). *Introduction to the Theory of Thermal Neutron Scattering.* Cambridge University Press.

Vacher, R. & Boyer, L. (1972). *Brillouin scattering: a tool for the measurement of elastic and photoelastic constants. Phys. Rev. B*, **6**, 639–673.

Van Hove, L. (1954). *Correlations in space and time and Born approximation scattering in systems of interacting particles. Phys. Rev.* **95**, 249–262.

Venkataraman, G., Feldkamp, L. A. & Sahni, V. C. (1975). *Dynamics of Perfect Crystals.* Cambridge: MIT Press.

Walker, C. B. (1956). *X-ray study of lattice vibrations in aluminum. Phys. Rev.* **103**, 547–557.

Williams, D. E. (1967). *Non-bonded potential parameters derived from crystalline hydrocarbons. J. Chem. Phys.* **47**, 4680–4684.

Willis, B. T. M. (1986). *Determination of the velocity of sound in crystals by time-of-flight neutron diffraction. Acta Cryst.* A**42**, 514–525.

Willis, B. T. M. & Pryor, A. W. (1975). *Thermal Vibrations in Crystallography.* Cambridge University Press.

International Tables for Crystallography (2010). Vol. B, Chapter 4.2, pp. 492–539.

4.2. Disorder diffuse scattering of X-rays and neutrons

By F. Frey, H. Boysen and H. Jagodzinski

4.2.1. Introduction

Diffuse scattering of X-rays, neutrons and other particles is an accompanying effect in all diffraction experiments aimed at structure analysis with the aid of elastic scattering. In this case, the momentum exchange of the scattered photon or particle includes the crystal as a whole; the energy transfer involved becomes negligibly small and need not be considered in diffraction theory. Static distortions as a consequence of structural changes cause typical elastic diffuse scattering. Many structural phenomena and processes contribute to diffuse scattering, and a general theory has to include all of them. Hence the exact treatment of diffuse scattering becomes very complex.

Inelastic scattering is due to dynamical fluctuations or ionization processes and may become observable as a 'diffuse' contribution in a diffraction pattern. A separation of elastic from inelastic diffuse scattering is generally possible, but difficulties may result from small energy exchanges that cannot be resolved for experimental reasons. The latter is true for scattering of X-rays by phonons, which have energies of the order of 10^{-2}–10^{-3} eV, values which are considerably smaller than 10 keV, a typical value for X-ray quanta. Another equivalent explanation, frequently forwarded in the literature, is the high speed of X-ray photons, such that the rather slow motion of atoms cannot be 'observed' by them during diffraction. Hence, all movements appear as static displacement waves of atoms, and temperature diffuse scattering is pseudo-elastic for X-rays. This is not true in the case of thermal neutrons, which have energies comparable to those of phonons. Phonon-related or thermal diffuse scattering is discussed separately in Chapter 4.1, *i.e.* the present chapter is mainly concerned with the elastic (or pseudo-elastic other than thermal) part of diffuse scattering. A particularly important aspect concerns diffuse scattering related to phase transitions, in particular the critical diffuse scattering observed at or close to the transition temperature. In simple cases, a satisfactory description may be given with the aid of a 'soft phonon', which freezes at the critical temperature, thus generating typical temperature-dependent diffuse scattering. If the geometry of the lattice is maintained during the transformation (*i.e.* there is no breakdown into crystallites of different cell geometry), the diffuse scattering is very similar to diffraction phenomena described in this chapter. Sometimes, however, very complicated interim stages (ordered or disordered) are observed, demanding a complicated theory for their full explanation (see, *e.g.*, Dorner & Comes, 1977).

Obviously, there is a close relationship between thermodynamics and diffuse scattering in disordered systems representing a stable or metastable thermal equilibrium. From the thermodynamical point of view, the system is then characterized by its grand partition function, which is intimately related to the correlation functions used in the interpretation of diffuse scattering. The latter is nothing other than a kind of 'partial partition function' where two atoms, or two cell occupancies, are fixed such that the sum of all partial partition functions represents the grand partition function. This fact yields the useful correlation between thermodynamics and diffuse scattering mentioned above, which may well be used for a determination of thermodynamical properties of the crystal. This important subject shall not be included here for the following reason: real three-dimensional crystals generally exhibit diffuse scattering by defects and/or disordering effects that are not in thermal equilibrium. They are created during crystal growth, or are frozen-in defects formed at higher temperatures. Hence a thermodynamical interpretation of diffraction data needs a careful study of diffuse scattering as a function of temperature or some other thermodynamical parameters. This can be done in very rare cases only, so the omission of this subject seems justified.

As shown in this chapter, electron-density fluctuations and distribution functions of defects play an important role in the complete interpretation of diffraction patterns. Both quantities may best be studied in the low-angle scattering range. Hence many problems cannot be solved without a detailed interpretation of low-angle diffraction (also called small-angle scattering).

Disorder phenomena in magnetic structures are also not specifically discussed here. Magnetic diffuse neutron scattering and special experimental techniques constitute a large subject by themselves. Many aspects, however, may be analysed along similar lines to those given here.

Glasses, liquids or liquid crystals show typical diffuse diffraction phenomena. Particle-size effects and strains have an important influence on the diffuse scattering. The same is true for dislocations and point defects such as interstitials or vacancies. These defects are mainly described by their strain field, which influences the intensities of sharp reflections like an artificial temperature factor: the Bragg peaks diminish in intensity while the diffuse scattering increases predominantly close to them. These phenomena are less important from a structural point of view, at least in the case of metals or other simple structures. This statement is true as long as the structure of the 'kernel' of defects may be neglected when compared with the influence of the strain field. Whether dislocations in more complicated structures meet this condition is not yet known.

Commensurate and incommensurate modulated structures and quasicrystals frequently show a typical diffuse scattering, a satisfactory explanation of which demands extensive experimental and theoretical study. A reliable structure determination becomes very difficult in cases where the interpretation of diffuse scattering has not been incorporated. Many erroneous structural conclusions have been published in the past. The solution of problems of this kind needs careful thermodynamical consideration as to whether a plausible explanation of the structural data can be given.

For all of the reasons mentioned above, this article cannot be complete. It is hoped, however, that it will provide a useful guide for those who need a full understanding of the crystal chemistry of a given structure.

The study of disorder in crystals by diffuse-scattering techniques can be performed with X-rays, neutrons or electrons. Each of these methods has its own advantages (and disadvantages) and they often can (or have to) be used in a complementary way (*cf.* Chapter 4.3 of this volume). Electron diffraction and microscopy are usually restricted to relatively small regions in space and thus supply information on a local scale, *i.e.* local defect structures. Moreover, electron-microscopy investigations are carried out on thin samples (films), where the disorder could be different from the bulk, and, in addition, could be affected by the high heat load deposited by the impinging electron beam. X-rays and neutrons sample larger crystal volumes and thus provide thermodynamically more important information on *averages* of the disorder. These methods are also better suited to the analysis of long-range correlated cooperative disorder phenomena. On the other hand, electron microscopy and diffraction often allow more direct access to disorder and can therefore provide valuable

Copyright © 2010 International Union of Crystallography

information about the underlying model, which can then be used for a successful interpretation of X-ray and neutron diffuse scattering. Basic aspects of electron diffraction and microscopy in structure determination are treated in Chapter 2.5 of this volume.

There is no comprehensive treatment of all aspects of diffuse scattering. Essential parts are treated in the textbooks by James (1954), Wilson (1962), Wooster (1962), Krivoglaz (1969, 1996*a*,*b*), Schwartz & Cohen (1977), Billinge & Thorpe (1998), Schweika (1998), Nield & Keen (2001), Fultz & Howe (2002), Frey (2003) and Welberry (2004); handbook articles have been written by Jagodzinski (1963, 1964*a*,*b*, 1987), Schulz (1982), Welberry (1985) and Moss *et al.* (2003), and a series of relevant papers has been collected by Collongues *et al.* (1977).

Finally, we mention that different symbols and 'languages' are used in the various diffraction methods. Quite a few of the new symbols in use are not really necessary, but some are caused by differences in the experimental techniques. For example, the neutron scattering length b may usually be equated with the atomic form factor f in X-ray diffraction. The differential cross section introduced in neutron diffraction represents the intensity scattered into an angular range dΩ and an energy range dE. The 'scattering law' in neutron work corresponds to the square of an (extended) structure factor; the 'static structure factor', a term used by neutron diffractionists, is nothing other than the conventional Patterson function. The complicated resolution functions in neutron work correspond to the well known Lorentz factors in X-ray diffraction. These have to be derived in order to include all the techniques used in diffuse-scattering work. In this article we try to preserve the most common symbols. In particular, the scattering vector will be denoted as $\mathbf{H}$, which is more commonly used in crystallography than $\mathbf{Q} = 2\pi\mathbf{H}$.

4.2.2. Basic scattering theory

4.2.2.1. General

Diffuse scattering results from deviations from the identity of translationally invariant scattering objects and from long-range correlations in space and time. Fluctuations of scattering amplitudes and/or phase shifts of the scattered wavetrains reduce the maximum capacity of interference (leading to Bragg reflections) and are responsible for the diffuse scattering, *i.e.* scattering parts that are not located in distinct spots in reciprocal space. Unfortunately, the terms 'coherent' and 'incoherent' scattering used in this context are not uniquely defined in the literature. Since all scattering processes are correlated in space and time, there is no incoherent scattering at all in its strict sense. A similar relationship exists for 'elastic' and 'inelastic' scattering. Here pure inelastic scattering would take place if the momentum and the energy were transferred to a single scatterer; on the other hand, an elastic scattering process would demand a uniform exchange of momentum and energy with the whole crystal. Obviously, both cases are idealized and the truth lies somewhere in between. In spite of this, many authors use the term 'incoherent' systematically for the diffuse scattering away from the Bragg peaks when all diffuse maxima or minima are due to structure factors of molecules or atoms only. Although this definition is unequivocal as such, it is physically incorrect. Other authors use the term 'coherent' for Bragg scattering only; all diffuse contributions are then called 'incoherent'. This definition is clear and unique since it considers space and time, but it does not differentiate between incoherent and inelastic. In the case of neutron scattering, both terms are essential and neither can be abandoned.

In neutron diffraction, the term 'incoherent scattering' refers to scattering by uncorrelated nuclear spin orientations or by a random distribution of isotopes of the same element. Hence another definition of 'incoherence' is proposed for scattering processes that are uncorrelated in space and time. In fact, there may be correlations between the spins *via* their magnetic field, but the correlation length in space (and time) may be very small, such that the scattering process appears to be incoherent. Even in these cases, the nuclei contribute to coherent (average structure) and incoherent scattering (diffuse background). Hence the scattering process cannot really be understood by assuming nuclei that scatter independently. For this reason, it seems to be useful to restrict the term 'incoherent' to cases where a random contribution to scattering is realized or, in other words, a continuous function exists in reciprocal space. This corresponds to a δ function in real four-dimensional space. The randomness may be attributed to a nucleus (neutron diffraction) or an atom (molecule). It follows from this definition that the scattering need not be constant, but may be modulated by structure factors of molecules. In this sense we shall use the term 'incoherent', remembering that it is incorrect from a physical point of view.

As mentioned in Chapter 4.1, the theory of thermal neutron scattering must be treated quantum mechanically. In principle, this is also true in the X-ray case. In the classical limit, however, the final expressions have a simple physical interpretation. Generally, the quantum-mechanical nature of the scattering function of thermal neutrons is negligible at higher temperatures and in those cases where energy or momentum transfers are not too large. In almost all disorder problems this classical interpretation is sufficient for the interpretation of diffuse-scattering phenomena. This is not quite true in the case of orientational disorder (plastic crystals) where H atoms are involved.

The basic formulae given below are valid in either the X-ray or the neutron case: the atomic form factor f replaces the coherent scattering length $b_{\rm coh}$ (abbreviated as b). The formulation in the frame of the van Hove correlation function $G(\mathbf{r}, t)$ (classical interpretation, coherent part) corresponds to a treatment by a four-dimensional Patterson function $P(\mathbf{r}, t)$.

The basic equations for the differential cross sections are

$$\frac{\mathrm{d}^2\sigma_{\rm coh}}{\mathrm{d}\Omega\ \mathrm{d}(\hbar\omega)} = N\frac{|\mathbf{k}|}{|\mathbf{k}_0|}\frac{\langle b\rangle^2}{2\pi\hbar}\int_{\mathbf{r}}\int_{t} G(\mathbf{r}, t) \times \exp\{2\pi i(\mathbf{H}\cdot\mathbf{r} - \nu t)\}\ \mathrm{d}\mathbf{r}\,\mathrm{d}t, \quad (4.2.2.1a)$$

$$\frac{\mathrm{d}^2\sigma_{\rm inc}}{\mathrm{d}\Omega\ \mathrm{d}(\hbar\omega)} = N\frac{|\mathbf{k}|}{|\mathbf{k}_0|}\frac{\langle b^2\rangle - \langle b\rangle^2}{2\pi\hbar}\int_{\mathbf{r}}\int_{t} G_{\rm s}(\mathbf{r}, t) \times \exp\{2\pi i(\mathbf{H}\cdot\mathbf{r} - \nu t)\}\ \mathrm{d}\mathbf{r}\,\mathrm{d}t, \quad (4.2.2.1b)$$

where N is the number of scattering nuclei of same chemical species; $\mathbf{k}$, $\mathbf{k}_0$ are the wavevectors after/before scattering and $\omega = 2\pi\nu$.

The integrations over space may be replaced by summations in disordered crystals, except in cases where structural elements exhibit a liquid-like behaviour. Then the van Hove correlation functions are

$$G(\mathbf{r}, t) = \frac{1}{N}\sum_{\mathbf{r}_j,\,\mathbf{r}_{j'}} \delta\{\mathbf{r} - [\mathbf{r}_{j'}(t) - \mathbf{r}_j(0)]\}, \quad (4.2.2.2a)$$

$$G_{\rm s}(\mathbf{r}, t) = \frac{1}{N}\sum_{\mathbf{r}_j} \delta\{\mathbf{r} - [\mathbf{r}_j(t) - \mathbf{r}_j(0)]\}. \quad (4.2.2.2b)$$

$G(\mathbf{r}, t)$ gives the probability that if there is an atom j at $\mathbf{r}_j(0)$ at time zero, there is an *arbitrary* atom j' at $\mathbf{r}_{j'}(t)$ at an *arbitrary* time t, while $G_{\rm s}(\mathbf{r}, t)$ refers to the *same* atom j at $\mathbf{r}_j(t)$ at time t.

Equations (4.2.2.1) may be rewritten using the four-dimensional Fourier transforms of G and $G_{\rm s}$, respectively:

$$S_{\rm coh}(\mathbf{H},\omega)=\frac{1}{2\pi}\int_{\mathbf{r}}\int_{t} G(\mathbf{r},t)\exp\{2\pi i(\mathbf{H}\cdot\mathbf{r}-\nu t)\}\,\mathrm{d}\mathbf{r}\,\mathrm{d}t, \quad (4.2.2.3a)$$

$$S_{\rm inc}(\mathbf{H},\omega)=\frac{1}{2\pi}\int_{\mathbf{r}}\int_{t} G_{\rm s}(\mathbf{r},t)\exp\{2\pi i(\mathbf{H}\cdot\mathbf{r}-\nu t)\}\,\mathrm{d}\mathbf{r}\,\mathrm{d}t, \quad (4.2.2.3b)$$

$$\frac{\mathrm{d}^2\sigma_{\rm coh}}{\mathrm{d}\Omega\,\mathrm{d}(\hbar\omega)}=N\frac{k}{k_0}\langle b\rangle^2 S_{\rm coh}(\mathbf{H},\omega), \quad (4.2.2.4a)$$

$$\frac{\mathrm{d}^2\sigma_{\rm inc}}{\mathrm{d}\Omega\,\mathrm{d}(\hbar\omega)}=N\frac{k}{k_0}[\langle b^2\rangle-\langle b\rangle^2]S_{\rm inc}(\mathbf{H},\omega). \quad (4.2.2.4b)$$

Incoherent scattering cross sections [(4.2.2.3*b*), (4.2.2.4*b*)] refer to one and the same particle (at different times). In particular, plastic crystals (see Section 4.2.5.5) may be studied by means of this incoherent scattering. It should be emphasized, however, that for reasons of intensity only disordered crystals with strong incoherent scatterers can be investigated by this technique. In practice, mostly samples that contain H atoms have been investigated. This topic will not be treated further in this article (see, *e.g.*, Springer, 1972; Lechner & Riekel, 1983). The following considerations are restricted to *coherent* scattering only.

Essentially the same formalism as given by equations (4.2.2.1*a*)–(4.2.2.4*a*) may be described using a generalized Patterson function, which is more familiar to crystallographers:

$$P(\mathbf{r},t)=\int_{\mathbf{r}'}\int_{t'=0}^{\tau}\rho(\mathbf{r}',t')\rho(\mathbf{r}'+\mathbf{r},t'+t)\,\mathrm{d}\mathbf{r}'\,\mathrm{d}t', \quad (4.2.2.5)$$

where ρ is the scattering density and τ denotes the time of observation. The only difference between $G(\mathbf{r},t)$ and $P(\mathbf{r},t)$ is the inclusion of the scattering weight (f or b) in $P(\mathbf{r},t)$. $P(\mathbf{r},t)$ is an extension of the usual spatial Patterson function $P(\mathbf{r})$. Similarly, S is replaced by another function,

$$\begin{aligned}2\pi S(\mathbf{H},\omega)&\equiv|F(\mathbf{H},\omega)|^2\\&=\int_{\mathbf{r}}\int_{t}P(\mathbf{r},t)\exp\{2\pi i(\mathbf{H}\cdot\mathbf{r}-\nu t)\}\,\mathrm{d}\mathbf{r}\,\mathrm{d}t,\end{aligned} \quad (4.2.2.6)$$

which is the Fourier transform of $P(\mathbf{r},t)$. One difficulty arises from neglecting the time of observation. Just as $S(\mathbf{H})$ [$\simeq|F(\mathbf{H})|^2$] is always proportional to the scattering volume V in the framework of kinematical theory or within Born's first approximation [*cf.* equation (4.2.2.1*a*)], so $S(\mathbf{H},\omega)$ [$\simeq|F(\mathbf{H},\omega)|^2$] is proportional to volume *and* observation time. In general, one does not make S proportional to τ, but one normalizes S to be independent of τ as $\tau\rightarrow\infty$: $2\pi S=(1/\tau)|F|^2$. Averaging over time τ gives therefore

$$\begin{aligned}S(\mathbf{H},\omega)=\frac{1}{2\pi}\int_{\mathbf{r}}\int_{t}\left\langle\int_{\mathbf{r}'}\rho(\mathbf{r}',t')\rho(\mathbf{r}'+\mathbf{r},t'+t)\,\mathrm{d}r'\right\rangle_{t'}\\\times\exp\{2\pi i(\mathbf{H}\cdot\mathbf{r}-\nu t)\}\,\mathrm{d}\mathbf{r}\,\mathrm{d}t.\end{aligned} \quad (4.2.2.7)$$

Special cases (see, *e.g.*, Cowley, 1981) are:

(1) *Pure elastic measurement*

$$\begin{aligned}I_{\rm e}\simeq S(\mathbf{H},0)&=\int_{\mathbf{r}}\left[\int_{t}P(\mathbf{r},t)\,\mathrm{d}t\right]\exp\{2\pi i\mathbf{H}\cdot\mathbf{r}\}\,\mathrm{d}\mathbf{r}\\&=\left|\sum_j f_j\langle\exp\{2\pi i\mathbf{H}\cdot\mathbf{r}_j(t)\}\rangle_t\right|^2.\end{aligned} \quad (4.2.2.8)$$

In this type of measurement, the time-averaged 'structure' is determined:

$$\langle\rho(\mathbf{r},t)\rangle_t=\int_{\mathbf{H}}|F(\mathbf{H},0)|\exp\{2\pi i\mathbf{H}\cdot\mathbf{r}\}\,\mathrm{d}\mathbf{H}.$$

The projection along the time axis in real (Patterson) space gives a section in Fourier space at $\omega=0$. True elastic measurement is a domain of neutron scattering. To determine the time-averaged structure of a statistically disordered crystal, dynamical disorder (phonon scattering) may be separated. For liquids or liquid-like systems this kind of scattering technique is rather ineffective, as the time-averaging procedure gives only a uniform particle distribution.

(2) *Integration over frequency (or energy)*

$$\begin{aligned}I_{\rm tot}\simeq\int_{\omega}|F(\mathbf{H},\omega)|^2\,\mathrm{d}\omega&=\int_{\omega}\int_{\mathbf{r}}\int_{t}P(\mathbf{r},t)\\&\times\exp\{2\pi i(\mathbf{H}\cdot\mathbf{r}-\nu t)\}\,\mathrm{d}\mathbf{r}\,\mathrm{d}t\,\mathrm{d}\omega\\&=\int_{\mathbf{r}}P(\mathbf{r},0)\exp\{2\pi i\mathbf{H}\cdot\mathbf{r}\}\,\mathrm{d}\mathbf{r}\end{aligned} \quad (4.2.2.9)$$

(*cf.* the properties of δ functions). In such an experiment, one determines the Patterson function for $t=0$, *i.e.* the instantaneous structure (a 'snapshot' of the correlation function): a projection in Fourier space along the energy axis gives a section in direct (Patterson) space at $t=0$. An energy integration is automatically performed in a conventional X-ray diffraction experiment ($|\mathbf{k}|\simeq|\mathbf{k}_0|$). One should bear in mind that in a real experiment there is, of course, an average over both the sample volume and the time of observation.

In most practical cases, averaging over time is equivalent to averaging over space: the total diffracted intensity may be regarded as the sum of the intensities from a large number of independent regions due to the limited coherence of a beam. At any time these regions take all possible configurations. Therefore, this sum of intensities is equivalent to the sum of intensities from any one region at different times,

$$\begin{aligned}\langle I_{\rm tot}\rangle_t&=\left\langle\sum_j\sum_{j'}f_jf_{j'}\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_j-\mathbf{r}_{j'})\}\right\rangle_t\\&=\sum_{j,j'}f_jf_{j'}\langle\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_j-\mathbf{r}_{j'})\}\rangle_t.\end{aligned} \quad (4.2.2.9a)$$

From the basic formulae one also derives the well known results for X-ray or neutron scattering by a periodic arrangement of particles in space [*cf.* equation (4.1.3.2) of Chapter 4.1]:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=N\frac{(2\pi)^3}{V_{\rm c}}\sum_{\mathbf{h}}|F(\mathbf{H})|^2\delta(\mathbf{H}-\mathbf{G}), \quad (4.2.2.10)$$

$$F(\mathbf{H})=\sum_j f_j(\mathbf{H})\exp\{-W_j\}\exp\{2\pi i\mathbf{H}\cdot\mathbf{r}_j\}. \quad (4.2.2.11)$$

$F(\mathbf{H})$ denotes the Fourier transform of one cell (structure factor); $\mathbf{G}$ is the reciprocal-lattice vector and W is the Debye–Waller factor; the f's are assumed to be real.

The evaluation of the intensity expressions (4.2.2.6), (4.2.2.8) or (4.2.2.9), (4.2.2.9*a*) for a disordered crystal must be performed in terms of statistical relationships between scattering factors and/or atomic positions.

4.2.2.2. X-ray and neutron scattering

From these basic scattering formulae some conclusions on the relative merits of X-rays and neutrons can be drawn. Both X-rays and thermal neutrons possess wavelengths of the order of interatomic distances and are thus well suited to the study of the atomic structure of condensed matter. Besides this, there are fundamental differences that make one or the other the radiation of choice for a particular problem or enable them to be used with advantage in a complementary manner. These differences are well documented in a number of textbooks on diffraction and are widely exploited in the determination of average structures. One fundamental difference is related to the interaction with matter: while X-rays are scattered by the electrons, neutrons are scat-

tered by the atomic nuclei (we are disregarding magnetic interactions in this chapter), *i.e.* X-rays probe the distribution of charges, which (in particular for light atoms) may not coincide with that of the nuclei. This fact forms the basis for the well known X − N technique in the analysis of Bragg intensities. No such studies of disorder diffuse scattering have been performed up to now, but may be a challenge for the future. As a consequence of the different size of the scattering objects, the scattering amplitude falls off with increasing scattering vector $|\mathbf{H}| = (\sin\theta)/\lambda$ for X-rays (the form factor), while it is constant for neutrons. Since for a complete interpretation of diffuse scattering high Q values are generally required, neutrons have their advantages in this respect. The scattering intensity varies as Z^2 (Z = number of electrons) for X-rays and more-or-less erratically for neutrons, for which it also depends on the specific isotope. This imposes problems for X-rays when trying to detect light elements (in particular H and O) in the presence of heavy elements or discriminating neighbouring elements in the periodic table (*e.g.* Al/Si/Mg, Fe/Co/Ni). On the other hand, these differences in scattering power can be used to identify the atomic species taking part in the disorder by comparing the intensity distributions observed with both methods. The contrast can be further enhanced by marking selected atoms by isotope substitution for neutrons or by using anomalous dispersion for X-rays.

In many practical cases, the complementarities of the two methods can be exploited by using them simultaneously. Some illustrative examples can be found in Boysen & Frey (1998).

Another important difference is in the energies given by the dispersion laws $E_x = \hbar ck$ and $E_n = (\hbar^2/2m)k^2$ for X-rays and neutrons, respectively, where c is the velocity of light, m is the neutron mass and $k = 2\pi/\lambda$ is the wavevector. For a typical wavelength, λ =1.54 Å (Cu $K\alpha$) ($k = 4.08$ Å^{-1}), $E_x = 8$ keV and E_n = 35 meV, *i.e.* the energies differ by almost six orders of magnitude. This means that neutron energies are similar to those of typical collective lattice excitations, which may therefore be determined more easily by neutron spectroscopy. The basic question of whether the underlying disorder is of static or dynamic origin can be answered using neutrons alone, *e.g.* by comparing the 'integral' (= elastic + inelastic) scattered intensity (*i.e.* without energy analysis, using a two-axis diffractometer) with the purely elastic intensity by placing an analyser crystal set to zero energy transfer in the diffracted beam (using a triple-axis diffractometer) or using time-of-flight methods. The high energy resolution of Mössbauer radiation can also be used for this purpose. Owing to the very low available intensities, however, this technique has only been applied occasionally. In the elastic mode, diffracted neutron intensities are also usually rather weak. The decision can, however, also be made simply by comparing the positions of the intensity maxima in reciprocal space. This follows from a consideration of the relative changes of the momentum transfer (the wavevector) in the two cases. For example, for an energy change $\Delta E = E_f - E_i$ (the subscripts refer to final and incident energies) of 1 meV, one obtains a change in the scattered wavevector Δk_f of 2×10^{-6} Å^{-1} for X-rays and 0.06 Å^{-1} for neutrons. In a two-axis experiment where the detector does not discriminate energies, *i.e.* without prior knowledge of the real energies, the distribution of diffuse intensities has to be drawn as if the scattering had been elastic ($\Delta E = 0$). In other words, the signal appears at a position displaced by Δk_f. From the estimations above it is clear that only in the neutron case can a measurable effect be expected.

Other specific features of neutrons include their magnetic moment (magnetic scattering is still a domain of neutrons, although progress is now being made with synchrotron radiation) and (nuclear) spin and isotope incoherent scattering processes (which allow the determination of the self-correlation functions). It is possible to separate the spin-incoherent part and distinguish between nuclear and magnetic scattering using polarization analysis. For most elements scattering and absorption is much weaker for neutrons, allowing larger samples to be analysed and sample environments (furnaces, pressure cells, electric and magnetic fields, special atmospheres *etc.*) to be handled much more easily. Differences in the experimental techniques for X-ray and neutron scattering are discussed in Section 4.2.8.

4.2.3. Qualitative treatment of structural disorder

4.2.3.1. Basic mathematics

Any structure analysis of disordered structures should start with a qualitative interpretation of diffuse scattering. This may be achieved with the aid of Fourier transforms and convolutions. A thorough mathematical treatment of Fourier transforms is given in Chapter 1.3 of this volume; here we give a simple short overview of the practical use of Fourier transforms, convolutions, their algebraic operations and examples of functions which are frequently used in diffraction physics (see, *e.g.*, Patterson, 1959; Jagodzinski, 1987). For simplicity, the following modified notation is used in this section: functions in real space are represented by lower-case letters, *e.g.* $a(\mathbf{r})$, $b(\mathbf{r})$, ... except for $F(\mathbf{r})$ [also, more frequently, denoted as $\rho(\mathbf{r})$] and $P(\mathbf{r})$, which are used as general symbols for a structure and the Patterson function, respectively; functions in reciprocal space are represented by capital letters $A(\mathbf{H})$, $B(\mathbf{H})$; and $\mathbf{r}$ and $\mathbf{H}$ are general vectors in real and reciprocal space, respectively. H, K, L denote continuous variables in reciprocal space; integer values are given by the commonly used symbols h, k, l. $Hx + Ky + Lz$ is the scalar product $\mathbf{H}\cdot\mathbf{r}$; d$\mathbf{r}$ and d$\mathbf{H}$ indicate integrations in three dimensions in real and reciprocal space, respectively. Even for X-rays, the electron density $\rho(\mathbf{r})$ will generally be replaced by the scattering potential $a(\mathbf{r})$. Consequently, anomalous contributions to scattering may be included if complex functions $a(\mathbf{r})$ are admitted. In the neutron case $a(\mathbf{r})$ refers to a quasi-potential. Using this notation we obtain the scattered amplitude

$$A(\mathbf{H}) = \int_{\mathbf{r}} a(\mathbf{r}) \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}\}\, \mathrm{d}\mathbf{r}, \qquad (4.2.3.1a)$$

$$a(\mathbf{r}) = \int_{\mathbf{H}} A(\mathbf{H}) \exp\{-2\pi i \mathbf{H}\cdot\mathbf{r}\}\, \mathrm{d}\mathbf{H} \qquad (4.2.3.1b)$$

(constant factors are omitted).

$a(\mathbf{r})$ and $A(\mathbf{H})$ are reversibly and uniquely determined by Fourier transformation. Consequently, equations (4.2.3.1) may simply be replaced by $a(\mathbf{r}) \leftrightarrow A(\mathbf{H})$, where the double-headed arrow represents the two integrations given by (4.2.3.1) and means: $A(\mathbf{H})$ is the Fourier transform of $a(\mathbf{r})$ and *vice versa.* The following relations may easily be derived from (4.2.3.1):

$$a(\mathbf{r}) + b(\mathbf{r}) \leftrightarrow A(\mathbf{H}) + B(\mathbf{H}) \quad \text{(law of addition)}, \qquad (4.2.3.2)$$

$$\beta a(\mathbf{r}) \leftrightarrow \beta A(\mathbf{H}) \quad \text{(law of scalar multiplication)}, \qquad (4.2.3.3)$$

where β is a scalar quantity.

On the other hand, the multiplication of two functions does not yield a relation of similar symmetrical simplicity:

$$a(\mathbf{r})b(\mathbf{r}) \leftrightarrow \int A(\mathbf{H}')B(\mathbf{H}-\mathbf{H}')\, \mathrm{d}\mathbf{H}' = A(\mathbf{H}) * B(\mathbf{H}), \qquad (4.2.3.4a)$$

$$a(\mathbf{r}) * b(\mathbf{r}) = \int a(\mathbf{r}')b(\mathbf{r}-\mathbf{r}')\, \mathrm{d}\mathbf{r}' \leftrightarrow A(\mathbf{H})B(\mathbf{H}) \qquad (4.2.3.4b)$$

(the laws of convolution and multiplication).

For simplicity, the complete convolution integral is abbreviated as $a(\mathbf{r}) * b(\mathbf{r})$. Since $a(\mathbf{r})b(\mathbf{r}) = b(\mathbf{r})a(\mathbf{r})$,

$$\int A(\mathbf{H}')B(\mathbf{H}-\mathbf{H}')\, \mathrm{d}\mathbf{H}' = \int B(\mathbf{H}')A(\mathbf{H}-\mathbf{H}')\, \mathrm{d}\mathbf{H}'$$

and *vice versa.* The convolution operation is commutative in either space.

The distribution law $a(b+c) = ab + ac$ is valid for the convolution as well:

$$a(\mathbf{r}) * [b(\mathbf{r}) + c(\mathbf{r})] = a(\mathbf{r}) * b(\mathbf{r}) + a(\mathbf{r}) * c(\mathbf{r}). \quad (4.2.3.5)$$

The associative law of multiplication does not hold if mixed products (convolution and multiplication) are used:

$$a(\mathbf{r}) * [b(\mathbf{r})c(\mathbf{r})] \neq [a(\mathbf{r}) * b(\mathbf{r})]c(\mathbf{r}). \quad (4.2.3.6)$$

From equations (4.2.3.1) one has

$$\begin{aligned} a(\mathbf{r} - \mathbf{r}_0) &\leftrightarrow A(\mathbf{H}) \exp\{2\pi i \mathbf{H} \cdot \mathbf{r}_0\}, \\ A(\mathbf{H} - \mathbf{H}_0) &\leftrightarrow a(\mathbf{r}) \exp\{-2\pi i \mathbf{H}_0 \cdot \mathbf{r}\} \end{aligned} \quad (4.2.3.7)$$

(the law of displacements).

Since symmetry operations are well known to crystallographers in reciprocal space as well, the law of inversion is only mentioned here:

$$a(-\mathbf{r}) \leftrightarrow A(-\mathbf{H}). \quad (4.2.3.8)$$

Consequently, if $a(\mathbf{r}) = a(-\mathbf{r})$, then $A(\mathbf{H}) = A(-\mathbf{H})$. In order to calculate the intensity, the complex conjugate $A^+(\mathbf{H})$ is needed:

$$a^+(\mathbf{r}) \leftrightarrow A^+(-\mathbf{H}), \quad (4.2.3.9a)$$

$$a^+(-\mathbf{r}) \leftrightarrow A^+(\mathbf{H}). \quad (4.2.3.9b)$$

Equations (4.2.3.9) yield the relationship $A^+(-\mathbf{H}) = A(\mathbf{H})$ ('Friedel's law') if $a(\mathbf{r})$ is a *real* function. The multiplication of a function with its conjugate is given by

$$a(\mathbf{r}) * a^+(-\mathbf{r}) \leftrightarrow |A(\mathbf{H})|^2,$$

with

$$a(\mathbf{r}) * a^+(-\mathbf{r}) = \int a(\mathbf{r}')a(\mathbf{r}' - \mathbf{r})\,\mathrm{d}\mathbf{r}' = P(\mathbf{r}). \quad (4.2.3.10)$$

Note that $P(\mathbf{r}) = P(-\mathbf{r})$ is not valid if $a(\mathbf{r})$ is complex. Consequently $|A(-\mathbf{H})|^2 \neq |A(\mathbf{H})|^2$. This is shown by evaluating $A(-\mathbf{H})A^+(-\mathbf{H})$,

$$A(-\mathbf{H})A^+(-\mathbf{H}) \leftrightarrow a(-\mathbf{r}) * a^+(\mathbf{r}) = P(-\mathbf{r}). \quad (4.2.3.11)$$

Equation (4.2.3.11) is very useful for the determination of the contribution of anomalous scattering to diffuse reflections.

Most of the diffuse-diffraction phenomena observed may be interpreted qualitatively or even semiquantitatively in a very simple manner using a limited number of important Fourier transforms, which are given below.

4.2.3.2. Fourier transforms

(1) *Normalized Gaussian function*

$$(\pi^{3/2}\alpha\beta\gamma)^{-1} \exp\{-(x/\alpha)^2 - (y/\beta)^2 - (z/\gamma)^2\}. \quad (4.2.3.12)$$

This plays an important role in statistics. Its Fourier transform is again a Gaussian:

$$\exp\{-\pi^2(\alpha^2H^2 + \beta^2K^2 + \gamma^2L^2)\}. \quad (4.2.3.12a)$$

The three parameters α, β, γ determine the width of the curve. Small values of α, β, γ represent a broad maximum in reciprocal space but a narrow one in real space and *vice versa*. The constant has been chosen such that the integral of the Gaussian is unity in real space. The product of two Gaussians in reciprocal space,

$$\begin{aligned} &\exp\{-\pi^2(\alpha_1^2H^2 + \beta_1^2K^2 + \gamma_1^2L^2)\} \\ &\quad \times \exp\{-\pi^2(\alpha_2^2H^2 + \beta_2^2K^2 + \gamma_2^2L^2)\} \\ &\quad = \exp\{-\pi^2[(\alpha_1^2 + \alpha_2^2)H^2 + (\beta_1^2 + \beta_2^2)K^2 \\ &\qquad + (\gamma_1^2 + \gamma_2^2)L^2]\} \end{aligned} \quad (4.2.3.12b)$$

again represents a Gaussian of the same type, but with a sharper profile. Consequently, its Fourier transform, which is given by the convolution of the transforms of the two Gaussians, is itself a Gaussian with a broader maximum. It may be concluded from this discussion that the Gaussian with $\alpha, \beta, \gamma \to 0$ is a δ function in real space and its Fourier transform is unity in reciprocal space.

(2) δ *'functions'*

For a proper definition see, *e.g.*, Cowley (1981). Here the term is used with the relaxed definition $\delta(\mathbf{r}) = 1$ if $\mathbf{r} = 0$ and zero elsewhere.

$$\delta(\mathbf{r}) \leftrightarrow 1 \text{ and } \delta(\mathbf{H}) \leftrightarrow 1$$

and

$$\delta(\mathbf{r} - \mathbf{r}_0) \leftrightarrow \exp\{2\pi i \mathbf{H} \cdot \mathbf{r}_0\}.$$

The convolution of two δ functions is again a δ function.

(3) *Lattices*

Lattices in real and reciprocal space may be described by δ functions:

$$l(\mathbf{r}) = \sum_{\mathbf{n}} \delta(\mathbf{r} - \mathbf{n})$$

and

$$L(\mathbf{H}) = \sum_{\mathbf{h}} \delta(\mathbf{H} - \mathbf{G}),$$

where $\mathbf{n} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$ and $\mathbf{G} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ represent the components of the translation vectors in real and reciprocal space, respectively. The Fourier transforms of lattices with orthogonal basis vectors of unit length and an infinite number of points in all three dimensions correspond to each other. In the following the relation $l(\mathbf{r}) \leftrightarrow L(\mathbf{H})$ is used in this generalized sense.

The Fourier transforms of *finite* lattices with N_1, N_2, N_3 nodes in the basic directions $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are given by

$$\frac{\sin \pi N_1 H}{\sin \pi H} \frac{\sin \pi N_2 K}{\sin \pi K} \frac{\sin \pi N_3 L}{\sin \pi L}, \quad (4.2.3.13)$$

which is a periodic function in reciprocal space, but, strictly speaking, nonperiodic in real space. It should be pointed out that the correspondence of lattices in either space is valid only if the origin coincides with a δ function. This fact may easily be understood by applying the law of displacement given in equation (4.2.3.7).

(4) *Box functions*

The Fourier transform of a box function $b(\mathbf{r})$ with unit height is

$$b(\mathbf{r}) \leftrightarrow \frac{\sin \pi\alpha H}{\pi H} \frac{\sin \pi\beta K}{\pi K} \frac{\sin \pi\gamma L}{\pi L}. \quad (4.2.3.14)$$

α, β, γ describe its extension in the three dimensions. This function is real as long as the centre of symmetry is placed at the origin, otherwise the law of displacement has to be used. The convolution of the box function with its inverse is needed for the calculation of intensities:

$$\begin{aligned} t(\mathbf{r}) &= b(\mathbf{r}) * b(-\mathbf{r}) \\ &\leftrightarrow \left(\frac{\sin \pi\alpha H}{\pi H}\right)^2 \left(\frac{\sin \pi\beta K}{\pi K}\right)^2 \left(\frac{\sin \pi\gamma L}{\pi L}\right)^2. \end{aligned} \quad (4.2.3.15)$$

$t(\mathbf{r})$ is a generalized three-dimensional 'pyramid' of doubled basal length when compared with the corresponding length of the box function. The top of the pyramid has a height given by the number of unit cells covered by the box function. Obviously, the box function generates a particle size in real space by multiplying the infinite lattice $l(\mathbf{r})$ by $b(\mathbf{r})$. Fourier transformation yields a particle-size effect well known in diffraction. Correspondingly, the termination effect of a Fourier synthesis is caused by multiplication by a box function in reciprocal space, which causes a broadening of maxima in real space.

(5) *Convolutions*

It is often very useful to elucidate the convolution given in equations (4.2.3.4) by introducing the corresponding pictures in real or reciprocal space. Since $1 \cdot f(\mathbf{r}) = f(\mathbf{r})$, $\delta(\mathbf{H}) * F(\mathbf{H}) = F(\mathbf{H})$, the convolution with a δ function must result in an identical picture of the second function, although the function is used as $f(-\mathbf{r})$ in the integrals of equations (4.2.3.4), $f(\mathbf{r} - \mathbf{r}')$ with $\mathbf{r}'$ as variable in the integral of convolution. The convolution with $f(-\mathbf{r})$ brings the integral into the form

$$\int f(\mathbf{r}')\, f(\mathbf{r}' - \mathbf{r})\, \mathrm{d}\mathbf{r}', \tag{4.2.3.16}$$

which is known as the Patterson (or self- or auto-convolution) function and represents the generalized Patterson function including anomalous scattering [*cf.* equation (4.2.3.10)].

The change of the variable in the convolution integral may sometimes lead to confusion if certain operations are applied to the arguments of the functions entering the integral. Hence it is useful to mention the invariance of the convolution integral with respect to a change of sign, or a displacement, respectively, if applied to $\mathbf{r}'$ in both functions. Consequently, the convolution with the inverted function $a(\mathbf{r}) * b(-\mathbf{r})$ may be determined as follows:

$$\begin{aligned} b'(\mathbf{r}) &= b(-\mathbf{r}), \\ a(\mathbf{r}) * b(-\mathbf{r}) &= a(\mathbf{r}) * b'(\mathbf{r}) = \int a(\mathbf{r}')b'(\mathbf{r} - \mathbf{r}')\, \mathrm{d}\mathbf{r}', \\ &= \int a(\mathbf{r}')b(\mathbf{r}' - \mathbf{r})\, \mathrm{d}\mathbf{r}' = P'(\mathbf{r}). \end{aligned} \tag{4.2.3.17}$$

This equation means that the second function is displaced into the positive direction by $\mathbf{r}$, then multiplied by the first function and integrated. In the original meaning of the convolution, the operation represents a displacement of the second function in the positive direction and an inversion at the displaced origin before multiplication and subsequent integration. On comparing the two operations it may be concluded that $P'(\mathbf{r}) \neq P'(-\mathbf{r})$ if the second function is acentric. For real functions both have to be acentric. In a similar way, it may be shown that the convolution

$$\begin{aligned} a(\mathbf{r} - \mathbf{m}) * b(\mathbf{r} - \mathbf{m}') &= \int_{\mathbf{r}'} a(\mathbf{r}' - \mathbf{m})b(\mathbf{r} - \mathbf{m}' - \mathbf{r}')\, \mathrm{d}\mathbf{r}' \\ &= \int_{\mathbf{r}''} a(\mathbf{r}'')b(\mathbf{r} - \mathbf{m}' - \mathbf{m} - \mathbf{r}'')\, \mathrm{d}\mathbf{r}''. \end{aligned} \tag{4.2.3.18}$$

Equation (4.2.3.18) indicates a displacement by $\mathbf{m}' + \mathbf{m}$ with respect to the convolution of the undisplaced functions. Consequently,

$$\delta(\mathbf{r} - \mathbf{m}) * \delta(\mathbf{r} - \mathbf{m}') = \delta(\mathbf{r} - \mathbf{m} - \mathbf{m}'). \tag{4.2.3.19}$$

Obviously, the commutative law of convolution is obeyed; on the other hand, the convolution with the inverted function yields

$$\delta(\mathbf{r} - \mathbf{m}' + \mathbf{m}),$$

indicating that the commutative law (interchange of $\mathbf{m}$ and $\mathbf{m}'$) is violated because of the different signs of $\mathbf{m}$ and $\mathbf{m}'$.

The effectiveness of the method outlined above may be greatly improved by introducing further Fourier transforms of useful functions in real and reciprocal space (Patterson, 1959).

4.2.3.3. General formulation of a disorder problem

From these basic concepts the generally adopted method in a disorder problem is to try to separate the scattering intensity into two parts, namely one part $\langle\rho\rangle$ from an average periodic structure where formulae (4.2.2.10), (4.2.2.11) apply and a second part $\Delta\rho$ resulting from fluctuations from this average (see, *e.g.*, Schwartz & Cohen, 1977). One may write this formally as

$$\rho = \langle\rho\rangle + \Delta\rho, \tag{4.2.3.20}$$

where $\langle\rho\rangle = \langle\rho_\mathrm{c}\rangle * l$ is defined to be time-independent and periodic in space and $\langle\Delta\rho\rangle = 0$. $\langle\rho_\mathrm{c}\rangle = (1/N)\sum_i \rho_i$ is the density of the average unit cell obtained by the projection of all unit cells into a single one. Fourier transformation gives the average amplitude,

$$\langle A\rangle = \langle F\rangle L = \sum_i \langle F\rangle \exp\{2\pi i\mathbf{H}\cdot\mathbf{r}_i\}, \tag{4.2.3.21a}$$

where $\langle F\rangle$ is the usual structure factor (4.2.2.11). The difference structure $\Delta\rho$ leads to the difference amplitude,

$$\begin{aligned} \Delta A = A - \langle A\rangle &= \sum_i (F_i - \langle F\rangle)\exp\{2\pi i\mathbf{H}\cdot\mathbf{r}_i\} \\ &= \sum_i \Delta F_i \exp\{2\pi i\mathbf{H}\cdot\mathbf{r}_i\}. \end{aligned} \tag{4.2.3.21b}$$

Because cross terms $\langle\rho\rangle * \Delta\rho$ vanish by definition, the Patterson function is

$$\begin{aligned} &[\langle\rho(\mathbf{r})\rangle + \Delta\rho(\mathbf{r})] * [\langle\rho(-\mathbf{r})\rangle + \Delta\rho(-\mathbf{r})] \\ &\quad = [\langle\rho(\mathbf{r})\rangle * \langle\rho(-\mathbf{r})\rangle] + [\Delta\rho(\mathbf{r}) * \Delta\rho(-\mathbf{r})]. \end{aligned} \tag{4.2.3.22}$$

Fourier transformation gives

$$I \simeq |\langle F\rangle|^2 + |\Delta F|^2, \tag{4.2.3.23a}$$

$$|\Delta F|^2 \simeq \langle |F|^2\rangle - |\langle F\rangle|^2. \tag{4.2.3.23b}$$

Since $\langle\rho\rangle$ is periodic, the first term in (4.2.3.23*a*) describes Bragg scattering,

$$I_\mathrm{B} = \langle A\rangle\langle A^*\rangle = |\langle F\rangle|^2 L, \tag{4.2.3.24}$$

where $\langle F\rangle$ plays the normal role of a structure factor of *one* cell of the averaged structure. The second term corresponds to diffuse scattering,

$$I_\mathrm{D} = \langle \Delta A \Delta A^*\rangle = \left\langle \sum_i \sum_j \Delta F_i \Delta F_j^* \exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_i - \mathbf{r}_j)\}\right\rangle. \tag{4.2.3.25}$$

In many cases, diffuse interferences are centred exactly at the positions of the Bragg reflections. It is then a serious experimental problem to decide whether the observed intensity distribution is due to Bragg scattering obscured by crystal-size limitations or due to other scattering phenomena.

If disordering is exclusively time-dependent, $\langle\rho\rangle$ represents the time average, whereas $\langle F\rangle$ gives the pure elastic scattering part [*cf.* (4.2.2.8)] and ΔF refers to inelastic scattering only.

4.2.3.4. General aspects of diffuse scattering

Diffuse scattering may be classified in various ways which may be related to specific aspects of the intensity distribution, *e.g.* according to the type of disorder: substitutional (or density or chemical) or displacive. The general expression (4.2.3.25), which may be rewritten as

$$I_{\rm D} = \left\langle \sum_k \left(\sum_i \Delta F_i \Delta F_{i+k} \right) \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}_k\} \right\rangle = \sum_k \langle \Delta F \Delta F^* \rangle_k \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}_k\}, \quad (4.2.3.26)$$

where $\mathbf{r}_k = \mathbf{r}_i - \mathbf{r}_j$, contains the difference structure factor ΔF, which may formally be written as

$$\Delta F = (\Delta f)\exp\{2\pi i \mathbf{H}\cdot(\mathbf{r}+\Delta\mathbf{r})\},$$

where (Δf) denotes a fluctuation of the scattering density (the form factor in the case of X-rays and the scattering length in the case of neutrons; note that this may include vacancies) and $\Delta\mathbf{r}$ denotes a fluctuation of the position, *i.e.* they refer to substitutional and displacive disorder, respectively. Although the two types often occur together in real crystals, they may be discriminated through their different dependence on the modulus of the scattering vector $\mathbf{H}$. This may be seen by considering the diffuse scattering of completely random fluctuations, *i.e.* without any correlations. For substitutional disorder one easily derives from (4.2.3.26)

$$I_{\rm D} = N\langle \Delta F^2 \rangle = N\langle \Delta f^2 \rangle,$$

while for displacive fluctuations with small amplitudes $\Delta\mathbf{r}$ the exponential may be expanded:

$$\exp\{2\pi i \mathbf{H}\cdot\Delta\mathbf{r}_i\} \simeq (1 + 2\pi i \mathbf{H}\cdot\Delta\mathbf{r}_i + \ldots),$$

leading to

$$I_{\rm D} = Nf^2 \sum_k \sum_i (2\pi\mathbf{H}\cdot\Delta\mathbf{r}_i)(2\pi\mathbf{H}\cdot\Delta\mathbf{r}_{i+k}) \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}_k\} \quad (4.2.3.27)$$

and

$$I_{\rm D} = Nf^2 4\pi^2 H^2 \langle \Delta r_i^2 \rangle_i. \quad (4.2.3.28)$$

Hence in the substitutional case the diffuse intensity is constant throughout reciprocal space, while in the displacive case it increases with H^2, *i.e.* it is zero near the origin. Any correlations will modulate this intensity, but will not change this general behaviour.

A second classification is related to the dimensionality of the disorder: one-dimensional disorder between two-dimensionally ordered objects (planes) leading to diffuse streaks perpendicular to the planes, two-dimensional disorder between one-dimensionally ordered objects (chains) leading to diffuse planes and general three-dimensional disorder.

These different types of disorder will be further discussed separately in the next paragraphs.

From the derivation in Section 4.2.3.3, one may note that information about an averaged disordered structure is contained in the Bragg scattering governed by

$$\langle F \rangle = \sum_i n_i f_i \exp\{2\pi\mathbf{H}\cdot\mathbf{r}_i\} T_i(\mathbf{H}), \quad (4.2.3.29)$$

where $T_i(\mathbf{H})$ is the Debye–Waller factor. Back-transformation to real space gives

$$\langle \rho \rangle = \sum_i n_i f_i \delta(\mathbf{r}-\mathbf{r}_i) * \text{p.d.f.}_i(\mathbf{r}-\mathbf{r}_i), \quad (4.2.3.30)$$

i.e. each average position is convoluted with p.d.f.$_i(\mathbf{r}-\mathbf{r}_i)$, the probability density function, which is the Fourier transform of $T_i(\mathbf{H})$. Any disorder model derived from the diffuse scattering must therefore comply with the p.d.f., *i.e.* this may and should be used to validate these models.

4.2.3.5. Types of diffuse scattering

4.2.3.5.1. Substitutional fluctuations, occupational disorder

As mentioned above, a completely random distribution of chemical species leads to a uniform distribution of diffuse intensity, which is also called monotonic Laue scattering. For example, for a binary alloy with scattering densities f_1 and f_2 and concentrations c_1 and c_2 this is simply given by $I_{\rm D} \propto c_1 c_2 (f_1 - f_2)^2$. Several authors use this, *i.e.* the intensity of a random distribution of occupancies, to define a so-called Laue unit, and therewith to put the general diffuse scattering on a relative scale.

Any deviations from the monotonic Laue scattering may be due either to the scattering factor of the objects (*e.g.* molecules) or to short-range-order correlations. For example, for a simple defect pair with distance $\mathbf{R}$ we have

$$I_{\rm D} \propto c_1 c_2 (f_1 - f_2)^2 [1 + \cos(2\pi\mathbf{H}\cdot\mathbf{R})], \quad (4.2.3.31a)$$

or for more general short-range order

$$I_{\rm D} \propto c_1 c_2 (f_1 - f_2)^2 \sum_n \alpha_n \cos(2\pi\mathbf{H}\cdot\mathbf{r}_n), \quad (4.2.3.31b)$$

where α_n are Warren–Cowley short-range-order parameters, as will be discussed in more detail in Section 4.2.5.4.

For a periodic modulation $\Delta f_n = f_A \cos(2\pi\mathbf{K}\cdot\mathbf{r}_n)$

$$I_{\rm D} \propto f_A^2 \left| \sum_n \left[\exp\{2\pi i(\mathbf{H}+\mathbf{K})\cdot\mathbf{r}_n\} + \exp\{2\pi i(\mathbf{H}-\mathbf{K})\cdot\mathbf{r}_n\} \right] \right|^2 \propto f_A^2 \sum_G \delta(\mathbf{H} - \mathbf{G} \pm \mathbf{K}), \quad (4.2.3.32)$$

where $\mathbf{G}$ is a reciprocal-lattice vector.

Hence a harmonic density modulation of a structure in real space leads to pairs of satellites in reciprocal space. Each main reflection is accompanied by a pair of satellites in the directions $\mp\Delta\mathbf{H}$ with phases $\mp 2\pi\varphi$. The reciprocal lattice may then be written in the following form:

$$L(\mathbf{H}) + \frac{\alpha}{2} L(\mathbf{H}+\Delta\mathbf{H}) \exp\{2\pi i \Phi\} + \frac{\alpha}{2} L(\mathbf{H}-\Delta\mathbf{H}) \exp\{2\pi i(-\Phi)\}, \quad (4.2.3.33)$$

where $0 \le \alpha \le 1$. Fourier transformation yields

$$l(\mathbf{r})\left[1 + \frac{\alpha}{2}\exp\{2\pi i(\Delta\mathbf{H}\cdot\mathbf{r}+\Phi)\} + \frac{\alpha}{2}\exp\{-2\pi i(\Delta\mathbf{H}\cdot\mathbf{r}+\Phi)\}\right] = l(\mathbf{r})[1 + \alpha\cos(2\pi\Delta\mathbf{H}\cdot\mathbf{r}+\Phi)]. \quad (4.2.3.34)$$

Equation (4.2.3.34) describes the lattice modulated by a harmonic density wave. Since phases cannot be determined by intensity measurements, there is no possibility of obtaining any information on the phase relative to the sublattice. From (4.2.3.34) it is obvious that the use of higher orders of harmonics does not change the situation. If $\Delta\mathbf{H}$ is not rational, such that no $n\Delta\mathbf{H}$ (n = integer) coincides with a main reflection in reciprocal space, the modulated structure is incommensurate with the basic lattice and the phase of the density wave becomes meaningless. The same is true for the relative phases of the various orders of harmonic modulations of the density. This uncertainty even remains valid for commensurate density modulations of the sublattice, because coinciding higher-order harmonics in reciprocal space cause the same difficulty; higher-order coefficients cannot uniquely be separated from lower ones, consequently structure determination becomes impossible unless phase-determination methods are applied. Fortunately, density modulations of pure harmonic character are impossible for chemical reasons; they may be approximated by disorder phenomena for the averaged structure only. If diffuse scattering is taken into

account, the situation is changed considerably: A careful study of the diffuse scattering alone, although difficult in principle, will yield the necessary information about the relative phases of density waves (Korekawa, 1967).

4.2.3.5.2. Displacement fluctuations, displacive disorder

Displacement modulations are more complicated, even in a primitive structure. The Fourier transform of a longitudinal or a transverse displacement wave has to be calculated and this procedure does not result in a function of similar simplicity. Formally, a periodic modulation $\Delta\mathbf{r}_n = \mathbf{a}\cos(2\pi\mathbf{K}\cdot\mathbf{r}_n)$ leads to

$$I_{\rm D} \propto (2\pi\mathbf{H}\cdot\mathbf{a})^2 \sum_{\mathbf{G},\nu} A_\nu \delta(\mathbf{H}-\mathbf{G}\pm\nu\mathbf{K}), \qquad (4.2.3.35)$$

where $\mathbf{a}$ is the amplitude of the displacement wave with $\mathbf{a}=\alpha\mathbf{e}$ and $\mathbf{e}$ is the polarization vector: $|\mathbf{e}|=1$. Equation (4.2.3.35) denotes a set of satellites whose amplitudes are described by Bessel functions of νth order, where ν represents the order of the satellites. The intensity of the satellites increases with the magnitude of the product $\mathbf{H}\cdot\mathbf{a}$. This means that a single harmonic displacement causes an infinite number of satellites. They may be unobservable at low diffraction angles as long as the amplitudes are small. If the displacement modulation is incommensurate there are no coincidences with reflections of the sublattice. Consequently, the reciprocal space is completely covered with an infinite number of satellites, or, in other words, with diffuse scattering. This is a clear indication that incommensurate displacement modulations belong to the category of disordered structures. Statistical fluctuations of amplitudes of the displacement waves cause additional diffuse scattering, regardless of whether the period is commensurate or incommensurate (Overhauser, 1971; Axe, 1980). Fluctuations of 'phases' (*i.e.* periods) cause a broadening of satellites in reciprocal space but no change in their integrated intensities as long as the changes are not correlated with fluctuation periods. The broadening of satellite reflections increases with the order of the satellites and $(\mathbf{H}\cdot\mathbf{a})$. Obviously, there is no fundamental difference in the calculation of diffuse scattering with an ordered supercell of sufficient size.

4.2.3.5.3. Clusters and domains

Many disorder problems may be treated qualitatively in terms of coarsened structures that are made up of clusters or domain-like order extending along one or more directions in space. Common crystals are made up of 'mosaic individuals' which are separated irregularly by unspecified defects such as dislocations, small-angle boundaries, micro-strain fields and other defects. This 'real' crystallinity is not covered by the term 'disorder'. Clustering is, in a structural sense, not a very well defined term, but refers to general agglomerations of atoms, vacancies, defects or atomic groups due to preferred chemical bonding or due to some kind of exsolution processes. In general, the term 'cluster' is used to describe some inhomogeneity in a basic matrix structure. The term 'domain' usually implies either a spatially varying structure forming separate blocks, such as occurring in twin domains, or a spatial variation of a physical property (*e.g.* magnetic moment), which may be visualized by different configurations (see, *e.g.*, Frey, 1997). Structural domains may be chemically homogeneous, as is the case in twin domain structures, or heterogeneous, which occurs, *e.g.*, in feldspars with their complicated Ca/Na and Al/Si distributions. A definition of a domain structure may be given by symmetry arguments or, equivalently, by the order-parameter concept. Individual domains of the coarsened structure may be derived from a, possibly hypothetical, high-symmetry aristophase obeying the concept of symmetry groups. A lower symmetry of the domain is either due to loss of a point-group symmetry element (twin domains) or due to loss of a translational symmetry element of the aristophase, giving rise to the formation of out-of-phase domains. The special case of antiphase domains is related to a violation of a translational vector $\mathbf{t}$ in the aristophase by regular or irregular insertion of lattice displacements $\frac{1}{2}\mathbf{t}$. Shear domains, which are related by other fractional parts of $\mathbf{t}$, may be explained by cooperative gliding of structural building blocks, for example coordination polyhedra. Colour (black-and-white) symmetry has to be used for magnetic domain structures and may also be used for chemical domain ordering. While preserving the same lattice, the disordered (usually high-temperature) phase, which is specified by grey, decomposes into black and white domains, possibly embedded in the grey matrix. However, the term 'atomic cluster' is frequently used in this context. Domains can exhibit a new order by themselves, thus creating new symmetries of the superstructure. The boundaries between different domains are apparently essential and may even be used for a definition of a domain. This does not mean that the boundaries or domain walls are simple atomic planes rather than extended intermediate structural states which mellow the transition from one domain to the next one. Thus extended walls may carry a 'gradient' structure between neighbouring differently oriented domains and may be treated as new domains with their own structure. If misfits at the planes of coincidence are accompanied by remarkable straining, an array of dislocations may destroy the exact symmetry relation between the individuals. There is a stepwise transition from fully coherent domains to fully incoherent crystal parts. As long as coherency between the individuals is preserved, domain structures can be treated simply by means of Fourier transforms and characteristic features of the disorder problem may be extracted from diffuse patterns.

Quite generally, the scattering density for a general arrangement of domains may be written as (Boysen, 1995; Frey, 1997)

$$[\rho_1 * l_1][b_1^1 * d_1^1 + b_1^2 * d_1^2 + \ldots] + [\rho_2 * l_2][b_2^1 * d_2^1 + b_2^2 * d_2^2 + \ldots] + \ldots, \qquad (4.2.3.36)$$

where $\rho_i = \rho_i(\mathbf{r})$ is the structure of the unit cell of domain i, $l_i(\mathbf{r})$ is the lattice function, $b_i^j(\mathbf{r})$ is the shape function (which is unity in the region of domain i with size j and zero elsewhere) and $d_i^j(\mathbf{r}) = \sum \delta(\mathbf{r}-\mathbf{m}_i^j)$ is the distribution function ($\mathbf{m}_i^j$ are the centres of the domains). Note again that domain walls may be included in this formulation as separate 'domains'. Fourier transformation yields the scattering amplitude

$$[F_1 L_1] * [B_1^1 D_1^1 + B_1^2 D_1^2 + \ldots] + [F_2 L_2] * [B_2^1 D_2^1 + B_2^2 D_2^2 + \ldots] + \ldots, \qquad (4.2.3.37)$$

where $F_i = F_i(\mathbf{H})$ is the usual structure factor, $L_i(\mathbf{H})$ is the lattice function in reciprocal space, $B_i^j(\mathbf{H})$ and $D_i^j(\mathbf{H})$ are the Fourier transforms of the shape and distribution function, respectively, and the intensity is

$$\sum_i \Big|[F_i L_i] * \Big[\sum_j B_i^j D_i^j\Big]\Big|^2 + 2\sum_{i\neq k}[F_i L_i] * \Big[\sum_j B_i^j D_i^j\Big][F_k^+ L_k^+] * \Big[\sum_j B_k^{j+} D_k^{j+}\Big]. \qquad (4.2.3.38)$$

The first term represents sharp or diffuse reflections that are modified by the convolution with the Fourier transforms of the shape and distribution function (more correctly, with the Fourier transforms of the corresponding Patterson functions), while the second (cross) term generally leads to smaller additional changes. It should be emphasized that this separation does not correspond to the separation into Bragg and diffuse scattering. Although a further mathematical treatment of this very general expression does not seem to be easy, some qualitative or even semi-quantitative conclusions may be drawn:

(1) If the d_i^j are coherent with l_i (*i.e.* the $\mathbf{m}_i^j$ coincide with lattice points $\sum m_i\mathbf{a}_i$), the convolution with the second term in (4.2.3.38) is meaningless and we obtain sharp (basic) Bragg

reflections. This means that domain sizes and distributions cannot be determined from the widths of the Bragg reflections. One has to rely on the second term in this case.

(*a*) If, in addition, the d_i^j are strictly periodic (which implies equal sizes b_i), superlattice reflections occur that accompany the basic reflections as *satellites*. If the periodicity is *commensurate* with the basic lattice, higher-order satellites may coincide with basic reflections. Hence care must be taken even in the determination of the average structure.

(*b*) If the periodicity in (*a*) is not perfect (there are small fluctuations of domain sizes), the satellites will be broadened.

(*c*) If the distribution of domain sizes is completely random, this will degenerate into a continuous intensity distribution (diffuse streaks). Note, however, that a completely random distribution cannot be realized, since the domains have a *finite* size, *i.e.* there will always be some kind of modulation of the streak.

(2) If the d_i^j are incoherent, the second terms vanish and the broadening effect of the convolution in the first term becomes effective, *i.e.* the domain size and distribution can be determined directly from the widths of the basic Bragg reflections. This situation may be encountered more frequently in real systems, as usually the fluctuating widths of the domain walls lead to a loss of coherence between different domains of the same type. Nevertheless, one has to be careful if relatively small domain sizes are present, since accidental coincidences may occur that lead to different broadening of different reflections. The problem of coherence effects in these cases is discussed and exemplified by Boysen (2007).

(3) Interferences between domains of *different* kinds may lead to specific extinction rules (governing the distribution of sharp and diffuse reflections). Some examples are given below.

(4) The general reciprocity between direct and reciprocal space means that narrow objects in real space correspond to broad features in reciprocal space and *vice versa*. In particular, for domain walls, which are usually thin in one direction and extended in the other two, we expect a broad feature in one direction (perpendicular to the wall) and narrow ones in the perpendicular directions, *i.e.* streaks. For a single layer (with an ideal interface) this streak will be uniform, *i.e.* without modulations. In order to estimate domain sizes W, the square of the Fourier transform of a box function (for a special direction $\mathbf{H}$)

$$B^2(H) = [\sin(\pi WH)/\pi H]^2 \qquad (4.2.3.39)$$

may be used to estimate the half widths of the peaks. Hence measuring the peak broadening in all directions of reciprocal space will give full information about the shape of the domains.

(5*a*) The situation becomes slightly more complicated when strain effects are present. For small extensions of the strained volume, *e.g.* inside a domain wall, it may be convenient to treat the whole wall as one unit cell, *i.e.* simply calculate the sum over all atoms directly. A strained lattice may generally be written

$$l_s(\mathbf{r}) = \sum_m \delta\big[\mathbf{r} - \textstyle\sum m_i \mathbf{a}_i - \mathbf{s}(m)\big], \qquad (4.2.3.40)$$

where the function $\mathbf{s}(m)$ describes the displacement variation in the different cells. Fourier transformation gives

$$L_s(\mathbf{H}) = \sum_m \exp\{2\pi i \mathbf{H} \cdot \textstyle\sum m_i \mathbf{a}_i\} \exp\{2\pi i \mathbf{H} \cdot \mathbf{s}(m)\}. \qquad (4.2.3.41)$$

For a harmonic (sinusoidal) wave $\mathbf{s} = \mathbf{z}_0 \sin(2\pi m/M + \varphi)$, where $\mathbf{z}_0$, φ and M are the amplitude, the phase and the period of the wave, one gets (Korekawa, 1967)

$$L_s(\mathbf{H}) = \sum_m J_m(2\pi \mathbf{H} \cdot \mathbf{z}_0) \exp\{2\pi i m\varphi\} L(H - m/M), \quad (4.2.3.41a)$$

i.e. a set of satellites m weighted by Bessel functions of order m. The intensity of the zeroth order (Bragg peak) thus decreases with $\mathbf{H}$, while higher orders become more important. Again, one has to be careful with possible coincidences. This dependence on the scattering vector $\mathbf{H}$ holds for all other functional forms of $\mathbf{s}(m)$ as well and is therewith fundamentally different to the behaviour of the size effect, which gives the same contribution to all Bragg reflections, like in the case of density fluctuations. This difference may therefore be used to distinguish between the two effects.

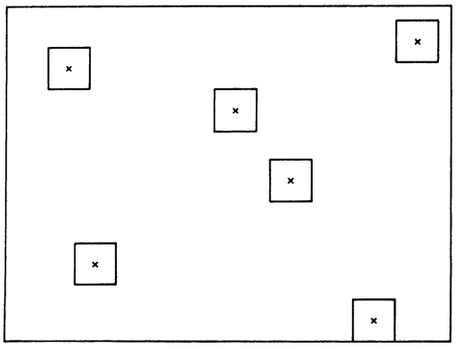


Fig. 4.2.3.1. Model of the two-dimensional distribution of point defects, causing changes in the surroundings.

(5*b*) Again, fluctuations of amplitude and phase lead to a broadening (diffuse tails) of the satellites. A general function $\mathbf{s}(m)$ may be analysed by expanding it into a Fourier series. For very long wavelengths (M) the satellites approach the basic Bragg peaks and cannot be observed easily. The case of a linear strain may be approximated by a sinusoidal modulation with very large wavelength, *i.e.* we will essentially observe a broadening of the Bragg peaks only. Then the strain ε may be estimated from the measured width $\Delta\mathbf{H}$:

$$\varepsilon = \Delta d/d = \Delta H/H.$$

Example: Clusters in a periodic lattice (low concentrations)

The exsolution of clusters of equal sizes is considered. The lattice of the host is undistorted, structure F_1, and the clusters have the same lattice but a different structure, F_2. A schematic drawing is shown in Fig. 4.2.3.1. Two different structures are introduced:

$$F_1(\mathbf{r}) = \sum_\nu \delta(\mathbf{r} - \mathbf{r}_\nu) * F_\nu(\mathbf{r}),$$

$$F_2(\mathbf{r}) = \sum_\mu \delta(\mathbf{r} - \mathbf{r}_\mu) * F_\mu(\mathbf{r}).$$

Their Fourier transforms are the structure factors $F_1(\mathbf{H})$, $F_2(\mathbf{H})$ of the matrix and the exsolved clusters, respectively. The boxes in Fig. 4.2.3.1 indicate the clusters, which may be represented by box functions $b(\mathbf{r})$ in the simplest case. It should be pointed out, however, that a more complicated shape means nothing other than a replacement of $b(\mathbf{r})$ by another shape function $b'(\mathbf{r})$ and its Fourier transform $B'(\mathbf{H})$. The distribution of clusters is represented by

$$d(\mathbf{r}) = \sum_{\mathbf{m}} \delta(\mathbf{r} - \mathbf{m}),$$

where $\mathbf{m}$ refers to the centres of the box functions (the crosses in Fig. 4.2.3.1). The problem is therefore defined by

$$l(\mathbf{r}) * F_1(\mathbf{r}) + [l(\mathbf{r})b(\mathbf{r})] * [F_2(\mathbf{r}) - F_1(\mathbf{r})] * d(\mathbf{r}). \qquad (4.2.3.42a)$$

The incorrect addition of $F_1(\mathbf{r})$ to the areas of clusters $F_2(\mathbf{r})$ is compensated by subtracting the same contribution from the second term in equation (4.2.3.42*a*). In order to determine the diffuse scattering, the Fourier transformation of (4.2.3.42*a*) is performed:

$$L(\mathbf{H})F_1(\mathbf{H}) + [L(\mathbf{H}) * B(\mathbf{H})][F_2(\mathbf{H}) - F_1(\mathbf{H})]D(\mathbf{H}). \quad (4.2.3.42b)$$

The intensity is given by

$$\left|L(\mathbf{H})F_1(\mathbf{H}) + [L(\mathbf{H}) * B(\mathbf{H})][F_2(\mathbf{H}) - F_1(\mathbf{H})]D(\mathbf{H})\right|^2. \quad (4.2.3.42c)$$

Evaluation of equation (4.2.3.42*c*) yields three terms (where c.c. means complex conjugate):

(i) $|L(\mathbf{H})F_1(\mathbf{H})|^2$

(ii) $\{[L(\mathbf{H})F_1(\mathbf{H})][L(\mathbf{H}) * B(\mathbf{H})] \times [F_2(\mathbf{H}) - F_1(\mathbf{H})]D(\mathbf{H}) + \text{c.c.}\}$

(iii) $|[L(\mathbf{H}) * B(\mathbf{H})][F_2(\mathbf{H}) - F_1(\mathbf{H})]D(\mathbf{H})|^2$.

The first two terms represent modulated lattices [multiplication of $L(\mathbf{H})$ by $F_1(\mathbf{H})$]. Consequently, they cannot contribute to diffuse scattering, which is completely determined by the third term. Fourier transformation of this term gives

$$\begin{aligned}&[l(\mathbf{r})b(\mathbf{r})] * \Delta F(\mathbf{r}) * d(\mathbf{r}) * [l(\mathbf{r})b(\mathbf{r})] * \Delta F^+(-\mathbf{r}) * d(-\mathbf{r})\\ &\quad = [l(\mathbf{r})b(\mathbf{r})] * [l(\mathbf{r})b(\mathbf{r})] * \Delta F(\mathbf{r}) * \Delta F^+(-\mathbf{r}) * d(\mathbf{r}) * d(-\mathbf{r})\\ &\quad = [l(\mathbf{r})t(\mathbf{r})] * \Delta F(\mathbf{r}) * \Delta F^+(-\mathbf{r}) * d(\mathbf{r}) * d(-\mathbf{r}), \quad (4.2.3.43a)\end{aligned}$$

where $l(\mathbf{r}) = l(-\mathbf{r})$, $b(\mathbf{r}) = b(-\mathbf{r})$ and $\Delta F = F_2 - F_1$. According to equation (4.2.3.15) and its subsequent discussion, the convolution of the two expressions in square brackets was replaced by $l(\mathbf{r})t(\mathbf{r})$, where $t(\mathbf{r})$ represents the 'pyramid' of n-fold height discussed above and n is the number of unit cells within $b(\mathbf{r})$. $d(\mathbf{r}) * d(-\mathbf{r})$ is the Patterson function of the distribution function $d(\mathbf{r})$. Its usefulness may be recognized by considering the two possible extreme solutions, namely the random and the strictly periodic distribution.

If no fluctuations of domain sizes are admitted, the minimum distance between two neighbouring domains is equal to the length of the domain in the corresponding direction. This means that the distribution function cannot be completely random. In one dimension, the solution of a random distribution of particles of a given size on a finite length shows that the distribution functions exhibit periodicities that depend on the average free volume of one particle (Zernike & Prins, 1927). Although the problem is more complicated in three dimensions, there should be no fundamental difference in the exact solutions.

On the other hand, it may be shown that the convolution of a pseudo-random distribution may be obtained if the average free volume is large. This is shown in Fig. 4.2.3.2(*a*) for the particular case of a cluster smaller than one unit cell. A strictly periodic distribution function (a superstructure) may result, however, if the volume of the domain and the average free volume are equal. Obviously, the practical solution for the self-convolution of the distribution function (which is the Patterson function) lies somewhere in between, as shown in Fig. 4.2.3.2(*b*). If a harmonic periodicity damped by a Gaussian is assumed, this self-convolution of the distribution in real space may be considered to consist of two parts, as shown in Figs. 4.2.3.2(*c*), (*d*). Note that the two different solutions result in completely different diffraction patterns:

(i) The geometrically perfect lattice extends to distances that are large when compared with the correlation length of the distribution function. Then the Patterson function of the distribution function concentrates at the positions of the basic lattice, which is given by multiplication by the lattice $l(\mathbf{r})$. The corresponding convolution in reciprocal space gives the same contribution to all Bragg reflections (Fig. 4.2.3.2*e*).

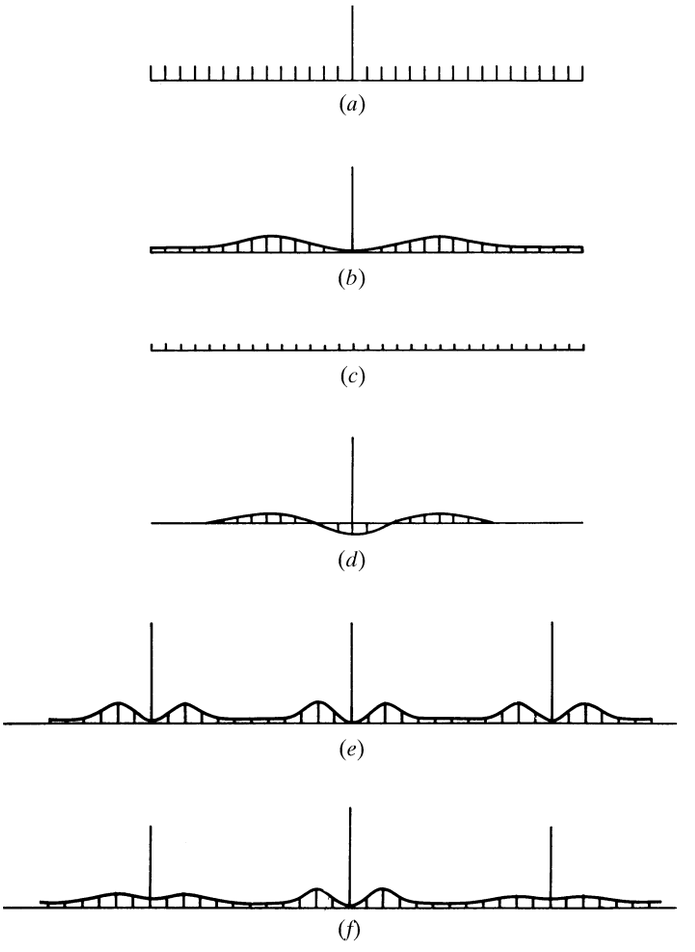


Fig. 4.2.3.2. One-dimensional Patterson functions of various point-defect distributions: (*a*) random distribution; (*b*) influence of finite volume of defects on the distribution function; (*c*), (*d*) decomposition of (*b*) into a periodic (*c*) and a convergent (*d*) part; (*e*) Fourier transform of (*c*) + (*d*); (*f*) changes of (*e*) if the centres of the defects show major deviations from the origins of the lattice.

(ii) There is no perfect lattice geometry. In this case, a continuous Patterson function results. Fourier transformation yields an influence that is now restricted primarily to the reflection 000, *i.e.* to the low-angle diffraction range.

Figs. 4.2.3.2(*e*), (*f*) show the different diffraction patterns of the diffuse scattering that is concentrated around the Bragg maxima. Although the discussion of the diffuse scattering was restricted to the case of identical domains, the introduction of a distribution of domain sizes does not influence the diffraction pattern essentially, as long as the fluctuation of sizes is small compared with the average volume of domain sizes and no strong correlation exists between domains of any size (a size-independent random distribution).

A complete qualitative discussion of the diffraction pattern may be carried out by investigating the Fourier transform of (4.2.3.43*a*):

$$[L(\mathbf{H}) * T(\mathbf{H})]|\Delta F(\mathbf{H})|^2|D(\mathbf{H})|^2. \quad (4.2.3.43b)$$

The first factor in (4.2.3.43*b*) describes the particle-size effect of a domain containing the influence of a surrounding strain field and the new structure of the domains precipitated from the bulk. $D(\mathbf{H})$ has its characteristic variation near the Bragg peaks (Figs. 4.2.3.2*e*,*f*) and is less important in between. For domain structure determination, intensities near the Bragg peaks should be avoided. Note that equation (4.2.3.43*b*) may be used for measurements using anomalous scattering in both the centric and the acentric case.

Solution of the diffraction problem. In equation (4.2.3.43*b*) $\Delta F(\mathbf{H})$ is replaced by its average,

$$\langle \Delta F(\mathbf{H}) \rangle = \sum_{\mu} p_{\mu} \Delta F_{\mu}(\mathbf{H}),$$

where p_{μ} represents the *a priori* probability of a domain of type μ. This replacement becomes increasingly important if small clusters (domains) have to be considered. Applications of the formulae to Guinier–Preston zones are given by Guinier (1942) and Gerold (1954); a similar application to clusters of vacancies in spinels with an excess of Al_2O_3 was outlined by Jagodzinski & Haefner (1967).

Although refinement procedures are possible in principle, the number of parameters entering the diffraction problem becomes increasingly large if more clusters or domains (of different sizes) have to be introduced. Another difficulty results from the large number of diffraction data which must be collected to perform a reliable structure determination. There is no need to calculate the first two terms in equation (4.2.3.42*c*) which contribute to the sharp Bragg peaks only, because their intensity is simply described by the averaged structure factor $|\langle F(\mathbf{H}) \rangle|^2$. These terms may therefore be replaced by

$$|L(\mathbf{H})|^2 |\langle F(\mathbf{H}) \rangle|^2$$

with

$$|\langle F(\mathbf{H}) \rangle|^2 = \Big| \sum_{\mu} p_{\mu} F_{\mu}(\mathbf{H}) \Big|^2, \qquad (4.2.3.43c)$$

where p_{μ} is the *a priori* probability of the structure factor $F_{\mu}(\mathbf{H})$. It should be emphasized here that (4.2.3.43*c*) is independent of the distribution function $d(\mathbf{r})$ or its Fourier transform $D(\mathbf{H})$.

While a number of solutions to the diffraction problem may be found in the literature for one-dimensional (1D) disorder (1D distribution functions) (see, *e.g.*, Jagodzinski, 1949*a*,*b*,*c*; Cowley, 1976*a*,*b*; Adlhart, 1981; Pflanz & Moritz, 1992), these become increasingly more complicated with increasing dimensionality. These types are therefore discussed separately in the following.

4.2.3.5.4. One-dimensional disorder

This lamellar (1D) type of disorder is common in many crystals for energetic reasons. For example, most of the important rock-forming minerals exhibit such disorder behaviour. Therefore we give some extended introduction to this field. To illustrate the application of the formalism outlined above, we start with a simple example of domains of an identical structure $\rho(\mathbf{r})$ displaced relative to each other by an arbitrary fault vector $\mathbf{\Delta}$ (Boysen *et al.*, 1991). If $t(\mathbf{r})$ describes the regions of one domain, then $1 - t(\mathbf{r})$ describes those of the displaced domains and the complete structure may be written

$$t(\mathbf{r})[l(\mathbf{r}) * \rho(\mathbf{r})] + [1 - t(\mathbf{r})][l(\mathbf{r} - \mathbf{\Delta}) * \rho(\mathbf{r})].$$

Introducing $t'(\mathbf{r}) = 2t(\mathbf{r}) - 1$, this may be rewritten as

$$[l(\mathbf{r}) + l(\mathbf{r} - \mathbf{\Delta})] * \rho(\mathbf{r})/2 = t'(\mathbf{r})\{[l(\mathbf{r}) - l(\mathbf{r} - \mathbf{\Delta})] * \rho(\mathbf{r})\}/2.$$

Fourier transformation yields

$$L(\mathbf{H})[1 + \exp\{2\pi i \mathbf{\Delta} \cdot \mathbf{r}\}]F(\mathbf{H})/2 + T'(\mathbf{H}) * L(\mathbf{H})[1 - \exp\{2\pi i \mathbf{\Delta} \cdot \mathbf{r}\}]F(\mathbf{H})/2.$$

The first term describes sharp reflections due to the multiplication with the lattice function $L(\mathbf{H})$, while the second term gives diffuse reflections due to the convolution with $T'(\mathbf{H})$. The corresponding intensities are

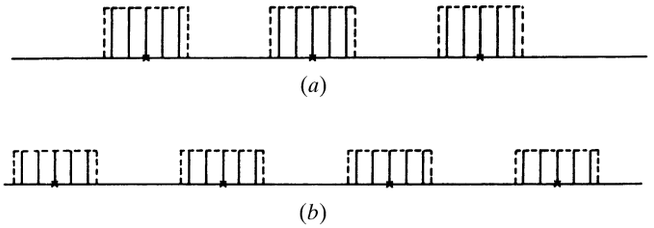


Fig. 4.2.3.3. Periodic array of domains consisting of two different atoms, represented by different heights. (*a*) Distribution of domain type 1, (*b*) distribution of domain type 2.

$$I_S \propto 1 + \cos(2\pi \mathbf{\Delta} \cdot \mathbf{H}) \text{ and } I_D \propto 1 - \cos(2\pi \mathbf{\Delta} \cdot \mathbf{H}).$$

In the general case, all reflections consist of a superposition of sharp and diffuse intensities. Note that the well known separation of sharp and diffuse intensities for antiphase domains is obtained if $\mathbf{\Delta}$ equals $\frac{1}{2}$ times a lattice vector.

(1) *Periodic distribution of lamellar domains*

Here $d(\mathbf{r})$ is one-dimensional and can easily be calculated: a periodic array of two types of lamellae having the same basic lattice $l(\mathbf{r})$ but different structures is shown in Fig. 4.2.3.3. The sizes of the two types of lamellae may be different. The structure of the first domain type is given by a convolution with $F_1(\mathbf{r})$ (Fig. 4.2.3.3*a*) and that of the second domain type by $F_2(\mathbf{r})$ (Fig. 4.2.3.3*b*). Introducing $\langle F(\mathbf{r}) \rangle$ and $\Delta F(\mathbf{r})$, the structure in real space is described by

$$\begin{aligned}[l(\mathbf{r})b_1(\mathbf{r})] * d(\mathbf{r}) * F_1(\mathbf{r}) + [l(\mathbf{r})b_2(\mathbf{r})] * d(\mathbf{r}) * F_2(\mathbf{r}) \\ = \{[l(\mathbf{r})b_1(\mathbf{r}) + l(\mathbf{r})b_2(\mathbf{r})] * d(\mathbf{r})\} * \langle F(\mathbf{r}) \rangle \\ + [l(\mathbf{r})b_1(\mathbf{r}) - l(\mathbf{r})b_2(\mathbf{r})] * d(\mathbf{r}) * \Delta F(\mathbf{r}). \qquad (4.2.3.44a)\end{aligned}$$

Obviously, the first term in curly brackets in equation (4.2.3.44*a*) is no more than $l(\mathbf{r})$ itself and $d(\mathbf{r})$ is strictly periodic. $b_1(\mathbf{r})$ and $b_2(\mathbf{r})$ are box functions, mutually displaced by $\pm(n_1 + n_2)/2$ unit cells in the stacking direction [n_1, n_2 are the numbers of cells covered by $b_1(\mathbf{r})$ and $b_2(\mathbf{r})$, respectively].

Fourier transformation of equation (4.2.3.44*a*) yields

$$L(\mathbf{H})\langle F(\mathbf{H}) \rangle + \{L(\mathbf{H}) * [B_1(\mathbf{H}) - B_2(\mathbf{H})]\} D(\mathbf{H}) \Delta F(\mathbf{H}). \qquad (4.2.3.44b)$$

The first term in equation (4.2.3.44*b*) gives the normal sharp reflections of the average structure, while the second describes superlattice reflections [sublattice $L_s(\mathbf{H}) = D(\mathbf{H})$ in reciprocal space], multiplied by $\Delta F(\mathbf{H})$ and another 'structure factor' generated by the convolution of the reciprocal lattice $L(\mathbf{h})$ with $[B_1(\mathbf{H}) - B_2(\mathbf{H})]$ (*cf.* Fig. 4.2.3.3). Since the centres of $b_1(\mathbf{r})$ and $b_2(\mathbf{r})$ are mutually displaced, the expression in square brackets includes extinctions if $b_1(\mathbf{r})$ and $b_2(\mathbf{r})$ represent boxes equal in size. These extinctions are discussed below. It should be pointed out that $L_s(\mathbf{H})$ and its Fourier transform $l_s(\mathbf{r})$ are commensurate with the basic lattice as long as no change of the translation vector at the interface of the lamellae occurs. Obviously, $L_s(\mathbf{H})$ becomes incommensurate in the general case of a slightly distorted interface. Considerations of this kind play an important role in the discussion of modulated structures.

No assumption has been made so far about the position of the interface. This point is meaningless only in the case of a strictly periodic array of domains (with no diffuse scattering). Therefore it seems to be convenient to introduce two basis vectors parallel to the interface in real space, which demand a new reciprocal vector perpendicular to them defined by $(\mathbf{a}' \times \mathbf{b}')/V'$, where $\mathbf{a}'$, $\mathbf{b}'$ are the new basis vectors and V' is the volume of the supercell. As long as the new basis vectors are commensurate with the original lattice, the direction of the new reciprocal vector $\mathbf{c}'^*$, perpendicular to $\mathbf{a}'$, $\mathbf{b}'$, passes through the Bragg points of the original reciprocal lattice and the reciprocal lattice of the superlattice

remains commensurate as long as V' is a multiple of V ($V' = mV$, m = integer). Since the direction of **c** is arbitrary to some extent, there is no clear rule about the assignment of superlattice reflections to the original Bragg peaks. This problem becomes very important if extinction rules of the basic lattice and the superlattice have to be described together.

Example

We consider a b.c.c. structure with two kinds of atoms (1, 2) with a strong tendency towards superstructure formation (CsCl-type ordering). According to equations (4.2.3.43*b*,*c*) and (4.2.3.44*b*), in the case of negligible short-range order we obtain the following expressions for sharp and diffuse scattering (where c = concentration):

$$I_S = |cf_1(\mathbf{H}) + (1-c)f_2(\mathbf{H})|^2 \quad \text{for } h+k+l = 2n,$$
$$I_D = c(1-c)|f_1(\mathbf{H}) - f_2(\mathbf{H})|^2 \quad \text{elsewhere.}$$

With increasing short-range order the sharp reflections remain essentially unaffected, while the diffuse ones concentrate into diffuse maxima at **h** with $h+k+l = 2n+1$. This process is treated more extensively below. As long as the domains exhibit no clear interface, it is useful to describe the ordering process with the two possible cell occupations of a pair of different atoms; then contributions of equal pairs may be neglected with increasing short-range order. Now the two configurations 1, 2 and 2, 1 may be given with the aid of the translation $\frac{1}{2}(\mathbf{a}+\mathbf{b}+\mathbf{c})$. Hence the two structure factors are

$$F_1 \text{ and } F_2 = F_1 \exp\{\pi i(h+k+l)\}.$$

Since the two structure factors occur with the same probability, the equations for sharp and diffuse reflections become

$$I_S = \tfrac{1}{4}|F_1(\mathbf{H})|^2[1 + \exp\{\pi i(h+k+l)\}]^2,$$
$$I_D = \tfrac{1}{4}|F_1(\mathbf{H})|^2[1 - \exp\{\pi i(h+k+l)\}]^2.$$

It is well recognized that no sharp reflections may occur for $h+k+l = 2n+1$, and the same holds for the diffuse scattering if $h+k+l = 2n$. This extinction rule for diffuse scattering is due to suppression of the contributions of equal pairs. The situation becomes different for lamellar structures. Let us first consider the case of lamellae parallel to (100). The ordered structure is formed by an alternating sequence of monoatomic layers consisting of atoms of types 1 or 2. Hence the interface between two neighbouring domains is a pair of equal layers 1, 1 or 2, 2, which are not equivalent. Each interface of type 1 (2) may be described by an inserted layer of type 1 (2) and the chemical composition differs from 1:1 if one type of interface is preferred. Since the contribution of equal pairs has been neglected in deriving the extinction rule of diffuse scattering (see above), this rule is no longer valid. Because of the lamellar structure the diffuse intensity is concentrated into streaks parallel to (00l). Starting from the diffuse maximum (010), the diffuse streak passes over the sharp reflection 011 to the next diffuse one 012 *etc.*, and the extinction rule is violated as long as one of the two interfaces is predominant. Therefore, the position of the interface determines the extinction rule in this orientation.

A completely different behaviour is observed for lamellae parallel to (110). This structure is described by a sequence of equal layers containing atoms 1 and 2. The interface between two domains (exchange of the two different atoms) is now nothing other than the displacement parallel to the layer of the original one in the ordered sequence. Calculation of the two structure factors would involve displacements $\pm\frac{1}{2}(\mathbf{a}-\mathbf{b})$. Starting from the diffuse reflection 001, the diffuse streak parallel to (HH0) passes through (111), (221), (331), ...; *i.e.* through diffuse reflections only. On the other hand, rows (HH2) going through (002), (112), (222), ... do not show any diffuse scattering. Hence we have a new extinction rule for diffuse scattering originating from the orientation of interfaces. This fact is rather important in structure determination. For various reasons, lamellar interfaces show a strong tendency towards a periodic arrangement. In diffraction the diffuse streak then concentrates into quite sharp superstructure reflections. These are not observed on those rows of the reciprocal lattice that are free from diffuse scattering. The same extinction law is not valid in the case of the (100) orientation of the interfaces. Summarizing, we may state that three types of extinction rules have to be considered:

(*a*) Normal extinctions for the average structure.

(*b*) Extinction of the difference structure factors for diffuse scattering.

(*c*) Extinctions caused by the ordering process itself.

(2) *Lamellar system with two different structures, where* $\langle F(\mathbf{H})\rangle$ *and* $\Delta F(\mathbf{H})$ *do not obey any systematic extinction law*

The convolution of the second term in equation (4.2.3.44*b*) (*cf.* Fig. 4.2.3.3) may be represented by a convolution of the Fourier transform of a box function $B_1(\mathbf{H})$ with the reciprocal superlattice. Since $B_1(\mathbf{H})$ is given by $\sin(\pi m_s H)/(\pi H)$, where m_s is the number of cells in the supercell, the reader might believe that the result of the convolution may easily be determined quantitatively: this assumption is not correct because of the slow convergence of $B_1(\mathbf{H})$. The systematic concidences of the maxima, or minima, of $B_1(\mathbf{H})$ with the points of the superlattice in the commensurate case cause considerable changes in intensities, especially in the case of a small domain thickness. For this reason, an accurate calculation of the amplitudes of the satellites is necessary (Jagodzinski & Penzkofer, 1981):

(*a*) Bragg peaks of the basic lattice

$$I \simeq |\langle F(\mathbf{H})\rangle|^2; \qquad (4.2.3.45a)$$

(*b*) satellites: $\nu = 2n$ (n = integer except 0)

$$I \simeq |2\sin\pi\nu C/[\sin\pi\nu/(N_1+N_2)]\Delta F(\mathbf{H})|^2; \qquad (4.2.3.45b)$$

(*c*) satellites: $\nu = 2n+1$

$$I \simeq |2\cos\pi\nu C/[\sin\pi\nu/(N_1+N_2)]\Delta F(\mathbf{H})|^2. \qquad (4.2.3.45c)$$

Here ν is the order of the satellites, $C = \frac{1}{2}(N_1 - N_2)/(N_1 + N_2)$, and N_1 and N_2 are the number of cells within $b_1(\mathbf{r})$ and $b_2(\mathbf{r})$, respectively.

Obviously, there is again a systematic extinction rule for even satellites if $N_1 = N_2$.

Equation (4.2.3.45*b*) indicates an increasing intensity of first even-order satellites with increasing C. Intensities of first even and odd orders become nearly equal if $N_2 \simeq \frac{1}{2}N_1$. Smaller values of N_2 result in a decrease of intensities of both even and odd orders (no satellites occur if $N_2 = 0$). The denominators in equations (4.2.3.45*b*,*c*) indicate a decrease in intensity with increasing order of the satellites. The quantitative behaviour of the intensities needs a more detailed discussion of the numerator in equations (4.2.3.45*b*,*c*) with increasing order of the satellites. Obviously, there are two kinds of extinction rules to be taken into account: systematic absences for the various orders of satellites, and the usual extinctions for $\langle F(\mathbf{H})\rangle$ and $\Delta F(\mathbf{H})$. Each has to be considered separately in order to arrive at reliable conclusions. This different behaviour of the superlattice reflections (satellites) and that of the basic lattice may be represented by a multidimensional group-theoretical representation, as has been shown by de Wolff (1974), Janner & Janssen (1980*a*,*b*), de Wolff *et al.* (1981) and others.

(3) *Nonperiodic system of lamellar domains*

Following the discussion of equations (4.2.3.43), one may conclude that the fluctuations of domain sizes cause a broadening of satellites if the periodic distribution function has to be

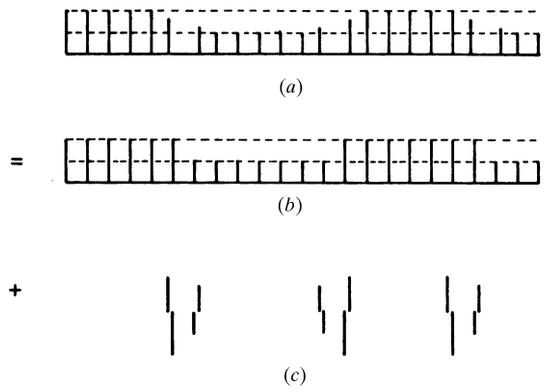


Fig. 4.2.3.4. Influence of distortions at the boundary of domains and separation into two parts; for discussion see text.

replaced by a statistical one. In this case, the broadening effect increases with the order of the satellites. The intensities, however, are completely determined by the distribution function and can be estimated by calculating the intensities of the perfectly ordered array, as approximated by the distribution function.

A careful check of $\langle F(\mathbf{H})\rangle$ and $\Delta F(\mathbf{H})$ in equations (4.2.3.45) shows that the position of the interface plays an important role in the intensities of the satellites. Since this position determines the origin of the unit cells in the sublattice, we have to choose this origin for the calculation of $F(\mathbf{H})$ and $\Delta F(\mathbf{H})$. This involves phase factors which are meaningless for integral values of $\mathbf{H}$ (i) if the average $\langle F(\mathbf{H})\rangle$ refers to different structures with arbitrary origin or (ii), which is important for practical cases, where no change occurs in the origin of related structures for neighbouring domains that are bound to an origin by general convention (*e.g.* a centre of symmetry). This statement is no longer true for non-integral values of $\mathbf{H}$, which are needed for the calculation of intensities of satellites. The intensities of satellites become different for different positions of the interface even in the absence of a relative displacement between neighbouring domains with respect to an origin by convention. This statement may be extended to nonperiodic distribution functions. Consequently, one may conclude that the study of diffuse scattering yields information on the interfacial scattering. For slightly different structures at the interface two cases are important:

(i) the two structures are related by symmetry (*e.g.* by a twin law); and

(ii) the difference between the two structures cannot be described by a symmetry operation.

In structures based on the same sublattice, the first case seems to be more important, because two different structures with the same sublattice are improbable. In the first case there is an identical sublattice if the symmetry operation in question does not influence the plane of intergrowth, *e.g.* a mirror plane should coincide with the plane of intergrowth. Since we have two inequivalent mirror planes in any sublattice, there are two such planes. It is assumed that no more than one unit cell of both domains at the interface has a slightly different structure without any change of geometry of the unit cell, and the number of unit cells is equal because of the equivalence of both domain structures (twins). Fig. 4.2.3.4(*a*) shows a picture of this model; Figs. 4.2.3.4(*b*), (*c*) explain that this structure may be described by two contributions:

(i) The first term is already given by equation (4.2.3.45) for $N_1 = N_2$, consequently only odd orders of satellites are observed.

(ii) The second term may be described by a superlattice containing $2N_1$ cells with an alternating arrangement of interfaces, correlated by the relevant plane of symmetry.

In real space, the second term may be constructed by convolution of the one-dimensional superlattice with two difference structures displaced by $\mp N_1/2$ units of the sublattice; its Fourier transformation yields

$$L_s(\mathbf{H})[\Delta F_i(\mathbf{H})\exp\{2\pi i N_1 H/2\} + \Delta F_i'(\mathbf{H})\exp\{-2\pi i N_1 H/2\}], \quad (4.2.3.46)$$

where ΔF_i, $\Delta F_i'$ correspond to the Fourier transforms of the contributions shown in Fig. 4.2.3.4(*c*). Since $H = \nu/N_1$, there are alternating contributions to the νth satellite which may be calculated more accurately by taking into account the symmetry operations. The important difference between equations (4.2.3.45) and (4.2.3.46) is the missing decrease in intensity with increasing order of the satellites. Consequently, one may conclude that the interface contributes to low- and high-order satellites as well, but its influence prevails for high-order satellites. Similar considerations may be made for two- and three-dimensional distributions of domains. A great variety of extinction rules may be found depending on the type of order approximated by the distribution under investigation.

(4) *Two kinds of lamellar domains with variable size distribution*

The preceding discussion of the diffuse scattering from domains is obviously restricted to relatively small fluctuations of domain sizes. This is specifically valid if the most probable domain size does not differ markedly from the average size. This condition is violated in the case of order–disorder phenomena. It may happen that the smallest ordered area is the most probable one, although the average is considerably larger. This may be shown for a lamellar structure of two types of layers correlated by a (conditional) pair probability $p_{\mu\mu'}(\mathbf{1})$. As shown below, a pair at distance $\mathbf{m}$ occurs with the probability $p_\mu p_{\mu\mu'}(\mathbf{m})$, which may be derived from the pair probability of nearest neighbours $p_\mu p_{\mu\mu'}(\mathbf{1})$. (In fact only one component of vector $\mathbf{m}$ is relevant in this context.) The problem will be restricted to two kinds of layers (μ, $\mu' = 1, 2$). Furthermore, it will be symmetric in the sense that the pair probabilities obey the following rules:

$$p_{11}(\mathbf{m}) = p_{22}(\mathbf{m}), \quad p_{12}(\mathbf{m}) = p_{21}(\mathbf{m}). \quad (4.2.3.47)$$

It may be derived from equation (4.2.3.47) that the *a priori* probabilities p_μ of a single layer are $\frac{1}{2}$ and

$$p_{11}(\mathbf{0}) = p_{22}(\mathbf{0}) = 1, \quad p_{12}(\mathbf{0}) = p_{21}(\mathbf{0}) = 0.$$

With these definitions and the general relation

$$p_{11}(\mathbf{m}) + p_{12}(\mathbf{m}) = p_{22}(\mathbf{m}) + p_{21}(\mathbf{m}) = 1,$$

the *a priori* probability of a domain containing m layers of type 1 may be calculated with the aid of $p_{11}(\mathbf{1})$ [$0 \le p_{11}(\mathbf{1}) \le 1$]:

$$p_\mu = \tfrac{1}{2}p_{11}(\mathbf{1})^{m-1}[1 - p_{11}(\mathbf{1})]. \quad (4.2.3.48)$$

Hence the most probable size of domains is a single layer, because a similar relation holds for layers of type 2. Since the average thickness of the domains is strongly dependent on $p_{11}(\mathbf{1})$ [infinite for $p_{11}(\mathbf{1}) = 1$ and one layer for $p_{11}(\mathbf{1}) = 0$], it may become very large in the latter case. Consequently, there are extremely large fluctuations if $p_{11}(\mathbf{1})$ is small but different from zero.

It may be concluded from equation (4.2.3.48) that the function $p_{11}(\mathbf{m})$ decreases monotonically with increasing $\mathbf{m}$, approaching $\frac{1}{2}$ with $\mathbf{m} \to \infty$. It is apparent that this cannot be true for a finite crystal if $p_{11}(\mathbf{m})$ is unity (a structure of two types of domains) or zero (a superstructure of alternating layers). In either case, the crystal should consist of a single domain of type 1 or 2, or one of the possible superstructures 1212 ..., 2121 ..., respectively. Hence one has to differentiate between long-range order, where two equivalent solutions have to be considered, and short-range order, where $p_{11}(\mathbf{m})$ approaches the *a priori* probability $\frac{1}{2}$ for large m. This behaviour of $p_{11}(\mathbf{m})$ and $p_{12}(\mathbf{m})$, which may also be

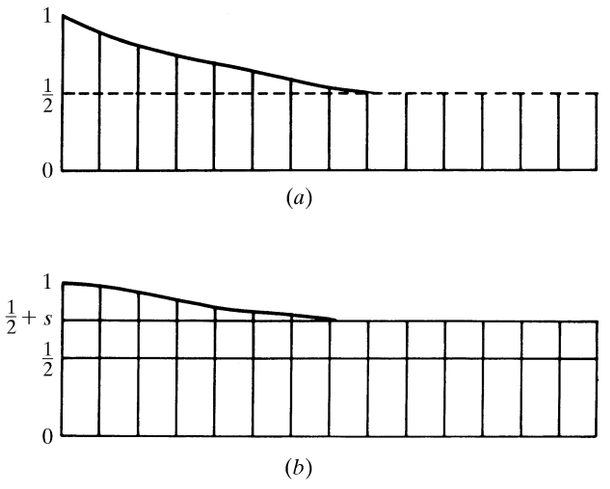


Fig. 4.2.3.5. Behaviour of $p_{11}(\mathbf{m})$ and $p_{12}(\mathbf{m})$ in mixed crystals (unmixing): (*a*) upper curve: short-range order only; (*b*) lower curve: long-range order.

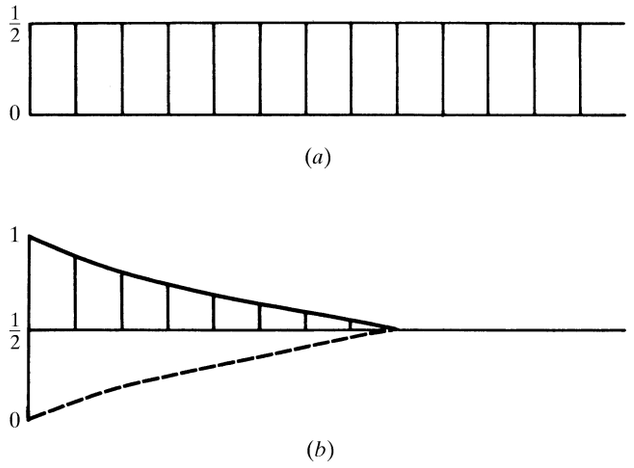


Fig. 4.2.3.6. Decomposition of Fig. 4.2.3.5(*a*) into (*a*) a periodic and (*b*) a rapidly convergent part.

expressed by equivalent correlation functions, is shown in Figs. 4.2.3.5(*a*) (short-range order) and 4.2.3.5(*b*) (long-range order, quantified by the parameter s). $p_{11}(\mathbf{m})$ approaches $\frac{1}{2}+s$ for large m with $s = 0$ in the case of short-range order, while $p_{12}(\mathbf{m})$ becomes $\frac{1}{2}-s$. Obviously, a strict correlation between $p_{11}(\mathbf{1})$ and s exists, and this has to be calculated. For a qualitative interpretation of the diffraction pictures this correlation may be derived from the diffraction pattern itself. The $p_{\mu\mu'}(\mathbf{m})$ are separable into a strictly periodic and a monotonically decreasing term approaching zero in both cases. This behaviour is shown in Figs. 4.2.3.6(*a*), (*b*). The periodic term contributes to sharp Bragg scattering. In the case of short-range order, the symmetry relations given in equation (4.2.3.47) are valid. The convolution in real space yields with factors $t(\mathbf{r})$ (equation 4.2.3.43*a*):

$$\begin{aligned}
&\tfrac{1}{2}t(\mathbf{r})\left[\sum_{\mathbf{m}}\delta(\mathbf{r}+\mathbf{m})p'_{11}(\mathbf{m})\right] * F_1(\mathbf{r}) * F_1(-\mathbf{r})\\
&\quad+\tfrac{1}{2}t(\mathbf{r})\left[\sum_{\mathbf{m}}\delta(\mathbf{r}+\mathbf{m})p'_{12}(\mathbf{m})\right] * F_1(\mathbf{r}) * F_2(-\mathbf{r})\\
&\quad+\tfrac{1}{2}t(\mathbf{r})\left[\sum_{\mathbf{m}}\delta(\mathbf{r}+\mathbf{m})p'_{21}(\mathbf{m})\right] * F_2(\mathbf{r}) * F_1(-\mathbf{r})\\
&\quad+\tfrac{1}{2}t(\mathbf{r})\left[\sum_{\mathbf{m}}\delta(\mathbf{r}+\mathbf{m})p'_{22}(\mathbf{m})\right] * F_2(\mathbf{r}) * F_2(-\mathbf{r}),
\end{aligned}$$

where $p'_{\mu\mu'}(\mathbf{m})$ are factors attached to the δ functions:

$$\begin{aligned}
p'_{11}(\mathbf{m}) &= p_{11}(\mathbf{m}) - \tfrac{1}{2} = p'_{22}(\mathbf{m}),\\
p'_{12}(\mathbf{m}) &= p'_{21}(\mathbf{m}) = -p'_{11}(\mathbf{m}).
\end{aligned}$$

The positive sign of $\mathbf{m}$ in the δ functions results from the convolution with the inverted lattice (*cf.* Patterson, 1959). Fourier transformation of the four terms given above yields the four corresponding expressions with $\mu, \mu' = 1, 2$:

$$\tfrac{1}{2}\left[T(\mathbf{H}) * \sum_{\mathbf{m}} p'_{\mu\mu'}(\mathbf{m})\exp\{-2\pi i\mathbf{H}\cdot\mathbf{m}\}\right]F_\mu(\mathbf{H})F^+_{\mu'}(\mathbf{H}). \qquad (4.2.3.49a)$$

Now the summation over $\mathbf{m}$ may be replaced by an integral if the factor $l(\mathbf{m})$ is added to $p'_{\mu\mu'}(\mathbf{m})$, which may then be considered as the smoothest continuous curve passing through the relevant integer values of $\mathbf{m}$:

$$\sum \rightarrow \int l(\mathbf{m})p'_{\mu\mu'}(\mathbf{m})\exp\{-2\pi i\mathbf{H}\cdot\mathbf{m}\}\,\mathrm{d}\mathbf{m}.$$

Since both $l(\mathbf{m})$ and $p'_{\mu\mu'}(\mathbf{m})$ are symmetric in our special case, we obtain

$$\sum = L(\mathbf{H}) * P'_{\mu\mu'}(\mathbf{H}).$$

Insertion of the sum in equation (4.2.3.49*a*) results in

$$\tfrac{1}{2}[L(\mathbf{H}) * T(\mathbf{H}) * P'_{\mu\mu'}(\mathbf{H})]F_\mu(\mathbf{H})F^+_{\mu'}(\mathbf{H}). \qquad (4.2.3.49b)$$

Using all symmetry relations for $p'_{\mu\mu'}(\mathbf{m})$ and $P'_{\mu\mu'}(\mathbf{H})$, respectively, we obtain for the diffuse scattering after summing over μ, μ'

$$I_{\mathrm{D}} \simeq [L(\mathbf{H}) * T(\mathbf{H}) * P'_{11}(\mathbf{H})]|\Delta F(\mathbf{H})|^2 \qquad (4.2.3.50)$$

with $\Delta F(\mathbf{H}) = \frac{1}{2}[F_1(\mathbf{H}) - F_2(\mathbf{H})]$.

It should be borne in mind that $P'_{11}(\mathbf{H})$ decreases rapidly if $p'_{11}(\mathbf{r})$ decreases slowly and *vice versa.* It is interesting to compare the different results from equations (4.2.3.43*b*) and (4.2.3.50). Equation (4.2.3.50) indicates diffuse maxima at the positions of the sharp Bragg peaks, while the multiplication by $D(\mathbf{H})$ causes satellite reflections in the neighbourhood of Bragg maxima. Both equations contain the factor $|\Delta F(\mathbf{H})|^2$, indicating the same influence of the two structures. More complicated formulae may be derived for several cell occupations. In principle, a result similar to equation (4.2.3.50) will be obtained, but more interdependent correlation functions $p_{\mu\mu'}(\mathbf{r})$ have to be introduced. Consequently, the behaviour of diffuse intensities becomes more differentiated in so far as all $p'_{\mu\mu'}(\mathbf{r})$ are now correlated with the corresponding $\Delta F_\mu(\mathbf{r})$, $\Delta F_{\mu'}(-\mathbf{r})$. Hence the method of correlation functions becomes increasingly ineffective with an increasing number of correlation functions. Here the cluster method seems to be more convenient and is discussed below.

(5) *Lamellar domains with long-range order: tendency to exsolution*

The Patterson function of a disordered crystal exhibiting long-range order is shown in Fig. 4.2.3.5(*b*). Now $p_{11}(\infty)$ converges against $\frac{1}{2}+s$ and the *a priori* probability changes correspondingly. Since $p_{12}(\infty)$ becomes $\frac{1}{2}-s$, the symmetry relation given in equation (4.2.3.47) is violated: $p_{11}(\mathbf{r}) \neq p_{22}(\mathbf{r})$ for a finite crystal; it is evident that another crystal shows long-range order with the inverted correlation function, $p_{22}(\infty) = \frac{1}{2}+s$, $p_{21}(\infty) = \frac{1}{2}-s$, such that the symmetry $p_{11}(\mathbf{r}) = p_{22}(\mathbf{r})$ is now valid for an assembly of finite crystals only. According to Fig. 4.2.3.5(*b*), there is a change in the intensities of the Bragg peaks.

$$\begin{aligned}
I_1 &\simeq |(\tfrac{1}{2}+s)F_1(\mathbf{H}) + (\tfrac{1}{2}-s)F_2(\mathbf{H})|^2,\\
I_2 &\simeq |(\tfrac{1}{2}+s)F_2(\mathbf{H}) + (\tfrac{1}{2}-s)F_1(\mathbf{H})|^2,
\end{aligned} \qquad (4.2.3.51)$$

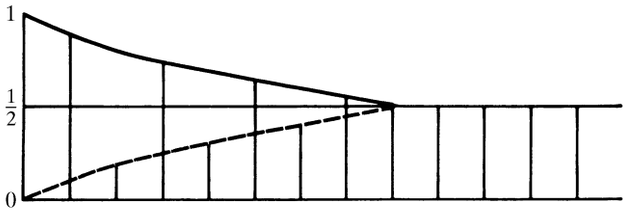


Fig. 4.2.3.7. The same distribution (*cf.* Fig. 4.2.3.5) in the case of superstructure formation.

where I_1, I_2 represent the two solutions discussed for the assembly of crystals and have to be added with the probability $\frac{1}{2}$; the intensities of sharp reflections become

$$I = (I_1 + I_2)/2. \tag{4.2.3.52}$$

Introducing equation (4.2.3.51) into (4.2.3.52), we obtain

$$I \simeq |\tfrac{1}{2}[F_1(\mathbf{H}) + F_2(\mathbf{H})]|^2 + s^2|F_1(\mathbf{H}) - F_2(\mathbf{H})^2|^2. \tag{4.2.3.53a}$$

$s = 0$ corresponds to the well known behaviour of sharp reflections, $s = \frac{1}{2}$ (maximum long-range order) gives

$$I \simeq \tfrac{1}{2}[|F_1(\mathbf{H})|^2 + |F_2(\mathbf{H})|^2]. \tag{4.2.3.53b}$$

This result reveals some difficulties for the determination of the averaged structure as long as s is different from zero or $\frac{1}{2}$, since in the former case the use of integrated sharp Bragg intensities yields a correct average structure. If $s = \frac{1}{2}$, a correct structure determination can only be performed with a refinement allowing for an incoherent superposition of two different structures. Having subtracted all periodic contributions to $p_{\mu\mu'}(\mathbf{r})$, new functions that describe the remaining nonperiodic parts have to be introduced (Fig. 4.2.3.6*b*). In order to obtain a clear overview of intensities, $p'_{\mu\mu'}(\mathbf{r})$ is again defined:

$$p'_{\mu\mu'}(\mathbf{r}) = cp_{\mu\mu'}(\mathbf{r}) - p_{\mu\mu'}(\infty),$$

where c should be chosen such that $p_{\mu\mu'}(\mathbf{0}) = 1$. By this definition a very simple behaviour of the diffuse scattering is obtained:

$$p'_{11}(\mathbf{r}) : \tfrac{1}{2} - s; \quad p'_{12}(\mathbf{r}) : -(\tfrac{1}{2} - s);$$
$$p'_{22}(\mathbf{r}) : \tfrac{1}{2} + s; \quad p'_{21}(\mathbf{r}) : -(\tfrac{1}{2} + s).$$

With the definitions introduced above it is found that

$$p'_{11}(\mathbf{r}) = p'_{22}(\mathbf{r}).$$

The diffuse scattering is given by

$$I_{1\mathrm{D}}(\mathbf{H}) = (\tfrac{1}{4} - s^2)|F_1(\mathbf{H}) - F_2(\mathbf{H})|^2[P'_{11}(\mathbf{H}) * L(\mathbf{H})]. \tag{4.2.3.54}$$

Since equation (4.2.3.54) is symmetrical with respect to an interchange of F_1 and F_2, the same result is obtained for $I_{2\mathrm{D}}$. Diffuse reflections occur in the positions of the sharp ones; the integrated intensities of sharp and diffuse reflections are independent of the special shape of $P'_{11}(\mathbf{H})$, hence

$$1 = \int P'_{11}(\mathbf{H}) \exp\{2\pi i \mathbf{0} \cdot \mathbf{H}\}\, \mathrm{d}\mathbf{H} = \int P'_{11}(\mathbf{H})\, \mathrm{d}\mathbf{H}$$

since $p_{11}(\mathbf{0}) = 1$.

(6) *Lamellar domains with long-range order: tendency to superorder*

So far it has been tacitly assumed that the crystal shows a preference for equal neighbours. If there is a reversed tendency (*i.e.* pairs of unequal neighbours are more probable), the whole procedure outlined above may be repeated as shown in Fig. 4.2.3.7 for the one-dimensional example. With the same probability of an unlike pair as used for the equal pair in the preceding example, the order process approaches an alternating structure such that the even-order neighbours have the same pair probabilities, while the odd ones are complementary for equal pairs (Fig. 4.2.3.7). In order to calculate intensities, it is necessary to introduce a new lattice with the doubled lattice constant and the corresponding reciprocal lattice with $b^{*\prime} = b^*/2$. In order to describe the probability $p_{\mu\mu'}(\mathbf{r})$, one has to introduce two lattices in real space – the normal lattice with the undisplaced origin and the displaced one. Fourier transformation of the new functions yields the following very similar results:

Sharp Bragg reflections

(*a*) $k' =$ even

$$I_{\mathrm{S}} = |\tfrac{1}{2}[F_1(\mathbf{H}) + F_2(\mathbf{H})]|^2. \tag{4.2.3.55a}$$

(*b*) $k' =$ odd

$$I_{\mathrm{S}} = s^2|\tfrac{1}{2}[F_1(\mathbf{H}) - F_2(\mathbf{H})]|^2. \tag{4.2.3.55b}$$

Diffuse reflections

(*c*) $k' =$ odd

$$I_{\mathrm{D}} = (\tfrac{1}{4} - s^2)|\tfrac{1}{2}[F_1(\mathbf{H}) - F_2(\mathbf{H})]|^2. \tag{4.2.3.55c}$$

In contrast to (5), this is the better situation for the determination of the averaged structure, which may be performed without any difficulty regardless of whether s is different from zero or not. For this purpose, even reflections (or reflections in the old setting) may be used. The inclusion of odd reflections in the structure determination of the superstructure is also possible if convenient **H**-independent scaling factors are introduced in order to compensate for the loss in intensity which is unavoidable for the integration of the diffuse scattering.

A few comments should be made on the physical meaning of the formulae derived above. All formulae may be applied to the general three-dimensional case, where long-range and short-range order is a function of the relevant thermodynamical parameters. In practice, long-range order will never be realized in a real crystal consisting of mosaic blocks which may behave as small subunits in order–disorder transitions. Another reason to assume partly incoherent areas in single crystals is the possible presence of strains or other distortions at the interfaces between domains, which should cause a decrease of the averaged areas of coherent scattering. All these effects may lead to diffuse scattering in the neighbourhood of Bragg peaks, similar to the diffuse scattering caused by domain structures. For this reason, an incoherent treatment of domains is probably more efficient, although considerable errors in intensity measurements may occur. A very careful study of line profiles is generally useful in order to decide between the various possibilities.

4.2.3.5.5. Two-dimensional disorder

The subject of two-dimensional disorder refers to predominantly one-dimensional structural elements, *e.g.* extended macromolecules and chain- or column-like structural units. A short introduction to this subject and some examples taken from inorganic structures are given in Section 4.2.5.3. Most important in this context, however, would be a treatment of disorder diffuse scattering of polymer/fibre structures. These are subjects in their own right and are treated in Chapter 4.5 of this volume and in Chapter 19.5 of Volume F.

4.2.3.5.6. Three-dimensional disorder

The solution of three-dimensional disorder problems is generally more demanding, although it may start with the formulation given above. Various algorithms have been developed to tackle these problems at least approximately, most of them restricted to particular models. Both real-space (cluster) and reciprocal-space (fluctuation wave) methods are employed and will be briefly addressed in Section 4.2.5.4. The more recently

developed approaches using computer simulations are described in Section 4.2.7.

Here we give only some general remarks on order–disorder problems.

Correlation functions in three dimensions may have very complicated periodicities; hence careful study is necessary to establish whether or not they may be interpreted in terms of a superlattice. If so, extinction rules have to be determined in order to obtain information on the superspace group. In the literature these are often called modulated structures (see Section 4.2.6) because a sublattice, as determined by the basic lattice, and a superlattice may well be defined in reciprocal space: reflections of a sublattice including (000) are formally described by a multiplication by a lattice having larger lattice constants (the superlattice) in reciprocal space; in real space this means a convolution with the Fourier transform of this lattice (the sublattice). In this way, the averaged structure is generated in each of the subcells (the superposition or 'projection' of all subcells into a single one). Obviously, the Patterson function of the averaged structure contains little information in the case of small subcells. Hence it is advisable to include the diffuse scattering of the superlattice reflections at the beginning of any structure determination.

N subcells in real space are assumed, each of them representing a kind of a complicated 'atom' that may be equal by translation or other symmetry operation. Once a superspace group has been determined, the usual extinction rules of space groups may be applied, remembering that the 'atoms' themselves may have systematic extinctions. Major difficulties arise from the existence of different symmetries of the subgroup and the supergroup. Since the symmetry of the supercell is lower in general, all missing symmetry elements may cause domains corresponding to the missing symmetry element: translations cause antiphase domains in their generalized sense; other symmetry elements cause twins generated by rotations, mirror planes or the centre of symmetry. If all these domains are small enough to be detected by a careful study of line profiles, using diffraction methods with a high resolution power, the structural study may be facilitated by introducing scaling factors for groups of reflections affected by the possible domain structures.

4.2.3.6. Symmetry

If disorder problems involving completely different structures (exsolutions *etc.*) are excluded, in general the symmetry of the diffuse-scattering pattern is the same as that of the Bragg peaks, *i.e.* it corresponds to the point group of the space group of the average structure. Only under specific directional growth conditions are deviations from this rule conceivable (although seemingly not very common). On the other hand, the symmetry of the underlying disorder model in direct space may be lower than that of the space group of the average structure (it is usually a subgroup of the space group of the average structure). The overall symmetry in reciprocal space is then restored by employing the missing (point-group) symmetry elements, which have therefore to be used in the calculation of the full diffraction pattern.

From these arguments, some specific disorder models may be classified according to the irreducible representations of the space group of the average structure. While for general wavevectors in the Brillouin zone no further restrictions appear, for high-symmetry directions consideration of the irreducible representations of the little co-group of the wavevector can help to identify the different symmetries of the disorder model. This becomes particularly evident when the modulation-wave approach is used as shown *e.g.* by Welberry & Withers (1990) and Welberry & Butler (1994). Of particular value are observed extinction rules, which may be calculated by group-theoretical methods as developed by Perez-Mato *et al.* (1998) for the extinctions occurring in inelastic neutron-scattering experiments. In favourable cases, the analysis of such extinctions alone can lead to a unique determination of the disorder model (see, *e.g.*, Aroyo *et al.*, 2002).

4.2.4. General guidelines for analysing a disorder problem

In general, the structure determination of a disordered crystal should start in the usual way by solving the average structure. The effectiveness of this procedure strongly depends on the distribution of integrated intensities of sharp and diffuse reflections. In cases where the integrated intensities of Bragg peaks are predominant, the maximum information can be drawn from the averaged structure. The observations of fractional occupations of lattice sites, split positions and anomalous large and anharmonic displacement parameters are indications of the disorder involved. Since these aspects of disorder phenomena in the averaged structure may be interpreted very easily, a detailed discussion of this matter is not given here (see any modern textbook of X-ray crystallography). Therefore, the anomalies of the average structure can give valuable hints on the underlying disorder and, *vice versa*, can be used to check the final disorder model derived from the diffuse scattering.

Difficulties may arise from the intensity integration, which should be carried out very carefully to separate the Bragg peaks from the diffuse contributions, *e.g.* by using a high-resolution diffraction method. The importance of this may be understood from the following argument. The averaged structure is determined by the coherent superposition of different structure factors. This interpretation is true if there is a strictly periodic subcell with long-range order that allows for a clear separation of sharp and diffuse scattering. There are important cases, however, where this procedure cannot be applied without loss of information.

(*a*) The diffuse scattering (other than thermal diffuse scattering) is concentrated near the Bragg peaks for a large number of reflections. Because of the limited resolution power of conventional single-crystal methods, the separation of sharp and diffuse scattering is impossible. Hence, the conventional study of integrated intensities does not really lead to an averaged structure. In this case, a refinement should be tried using an incoherent superposition of different structure factors (from the average structure and the difference structure). Application of this procedure is subject to conditions which have to be checked very carefully before starting the refinement: first, it is necessary to estimate the amount of diffuse scattering not covered by intensity integration of the 'sharp' reflections. Since loss in intensity, hidden in the background scattering, is underestimated very frequently, it should be checked whether nearly coinciding sharp and diffuse maxima are modulated by the same structure factor. It may be difficult to meet this condition in some cases; *e.g.* this condition is fulfilled for antiphase domains but the same is not true for twin domains.

(*b*) The concentration of diffuse maxima near Bragg peaks is normally restricted to domain structures with a strictly periodic sublattice. Cases deviating from this rule are possible. Since they are rare, they are omitted here. Even structures with small deviations from the average structure do not necessarily lead to structure factors for diffuse scattering that are proportional to those of the average structure. This has been shown in the case of a twin structure correlated by a mirror plane, where the reflections of a zone only have equal structure factors (Cowley & Au, 1978). This effect causes even more difficulties for orthogonal lattices, where the two twins have reflections in exactly the same positions, although differing in their structure factors. In this particular case, the incoherent or coherent treatment in refinements may be seriously hampered by strains originating from the boundary. Unsatisfactory refinements may be explained in this way but this does not improve their reliability.

The integrated intensity within a Brillouin zone of any structure is independent of atomic positions if the atomic form factors remain unchanged by structural fluctuations. Small deviations of atomic form factors owing to electron-density changes of valence electrons are neglected. Consequently, the integrated diffuse intensities remain unchanged if the average structure is not altered by the degree of order. The latter condition is obeyed in cases where a geometrical long-range order of the lattice is independent of the degree of order, and no long-range order in the structure exists. This law is extremely useful for the interpretation of diffuse scattering. Unfortunately, intensity integration over coinciding sharp and diffuse maxima does not necessarily lead to a structure determination of the corresponding undistorted structure. This integration may be useful for antiphase domains without major structural changes at the boundaries. In all other cases, the deviations of domains (or clusters) from the averaged structure determine the intensities of maxima, which are no longer correlated with those of the average structure.

If the integrated intensity of diffuse scattering is comparable with, or even larger than, those of the Bragg peaks, it is useful to begin the interpretation with a careful statistical study of the diffuse intensities. Intensity statistics can be applied in a way similar to the intensity statistics in classical structure determination. The following rules are briefly discussed in order to enable a semiquantitative interpretation of the essential features of disorder.

(1) First, it is recommended that the integrated intensities are studied in certain areas of reciprocal space.

(2) Since low-angle scattering is very sensitive to fluctuations of densities, the most important information can be drawn from its intensity behaviour. If there is at least a one-dimensional sublattice in reciprocal space without diffuse scattering, it may often be concluded that there is no important low-angle scattering either. This law is subject to the condition of a sufficient number of reflections obeying this extinction rule without any exception.

(3) If the diffuse scattering shows maxima and minima, it should be checked whether the maxima observed may be approximately assigned to a lattice in reciprocal space. Obviously, this condition can hardly be met exactly if these maxima are modulated by a kind of structure factor, which causes displacements of maxima proportional to the gradient of this structure factor. Hence this influence may well be estimated from a careful study of the complete diffuse diffraction pattern.

It should then be checked whether the corresponding lattice represents a sub- or a superlattice of the structure. An increase of the width of reflections as a function of increasing $|\mathbf{H}|$ indicates strained clusters of this sub- or superlattice.

(4) The next step is to search for extinction rules for the diffuse scattering. The simplest is the lack of low-angle scattering, which has already been mentioned above. Since diffuse scattering is generally given by equation (4.2.3.23*b*),

$$I_{\rm D}(\mathbf{H}) = \langle |F(\mathbf{H})|^2 \rangle - |\langle F(\mathbf{H}) \rangle|^2 = \sum_\mu p_\mu |F_\mu(\mathbf{H})|^2 - \Big| \sum_\mu p_\mu F_\mu(\mathbf{H}) \Big|^2,$$

it may be concluded that this condition is fulfilled in cases where all structural elements participating in disorder differ by translations only (stacking faults, antiphase domains *etc.*). They add phase factors to the various structure factors, which may become $n2\pi$ (n = integer) for specific values of the reciprocal vector $\mathbf{H}$. If all p_μ are equivalent by symmetry

$$p \sum_\mu |F_\mu(\mathbf{H})|^2 - \left[p \sum_\mu F_\mu(\mathbf{H}) \right] \left[p \sum_\mu F_\mu^+(\mathbf{H}) \right] = 0.$$

Other possibilities for vanishing diffuse scattering may be derived in a similar manner for special reflections if glide operations are responsible for disorder. Since we are concerned with disordered structures, these glide operations need not necessarily be a symmetry operation of the lattice. It should be pointed out, however, that all these extinction rules of diffuse scattering are a kind of 'anti'-extinction rule, because they are valid for reflections having maximum intensity for the sharp reflections unless the structure factor itself vanishes.

(5) Furthermore, it is important to plot the integrated intensities of sharp and diffuse scattering as a function of the reciprocal coordinates, at least in a semiquantitative way. If the ratio of integrated intensities remains constant in the statistical sense, we are predominantly concerned with a density phenomenon. It should be pointed out, however, that a particle-size effect of domains behaves like a density phenomenon (the density changes at the boundary!).

If the ratio of 'diffuse' to 'sharp' intensities increases with diffraction angle, we have to take into account atomic displacements. A careful study of this ratio yields very important information on the number of displaced atoms. The result has to be discussed separately for domain structures if the displacements are equal in the subcells of a single domain but different for the various domains. In the case of two domains with displacements of all atoms, the integrated intensities of sharp and diffuse reflections become statistically equal for large $|\mathbf{H}|$. Other rules may be derived from statistical considerations.

(6) The next step of a semiquantitative interpretation is to check the intensity distribution of diffuse reflections in reciprocal space. In general this modulation is simpler than that of the sharp reflections. Hence it is frequently possible to start a structure determination with diffuse scattering. This method is extremely helpful for one- and two-dimensional disorder where partial structure determinations yield valuable information, even for the evaluation of the average structure.

(7) In cases where no sub- or superlattice belonging to the diffuse scattering can be determined, a careful check of integrated intensities in the neighbourhood of Bragg peaks should again be performed. If systematic absences are found, the disorder is most probably restricted to specific lattice sites which may also be found in the average structure. The accuracy, however, is much lower here. If no such effects correlated with the average structure are observed, the disorder problem is related to a distribution of molecules or clusters with a structure differing from the average structure. As pointed out in Section 4.2.3.1, the problem of the representative structure(s) of the molecule(s) or the cluster(s) should be solved. Furthermore, their distribution function(s) is (are) needed. In this particular case, it is very useful to start with a study of diffuse intensity at low diffraction angles in order to acquire information about density effects. Despite the contribution to sharp reflections, one should remember that the level of information derived from the average structure may be very low (*e.g.* small displacements, low concentrations *etc.*).

(8) A Patterson picture – or strictly speaking a difference Patterson ($|\Delta F|^2$ Fourier synthesis) – may be very useful in this case. This method is promising in the case of disorder in molecular structures where the molecules concerned are at least partly known. Hence the interpretation of the difference Patterson may start with some internal molecular distances. Nonmolecular structures show some distances of the average structure. Consequently, a study of the important distances will yield information on displacements or replacements in the average structure. For a detailed study of this matter the reader is referred to the literature (Schwartz & Cohen, 1977).

Although it is highly improbable that exactly the same diffraction picture will be found, the use of an atlas of optical transforms (Wooster, 1962; Harburn *et al.*, 1975; Welberry & Withers, 1987) may be helpful at the beginning of any study of

diffuse scattering. Alternatively, computer simulations may be helpful, as discussed in Section 4.2.7. The most important step is the separation of the distribution function from the molecular scattering. Since this information may be derived from a careful comparison of low-angle diffraction with the remaining sharp reflections, this task is not too difficult. If the influence of the distribution function is unknown, the reader is strongly advised to disregard the immediate neighbourhood of Bragg peaks in the first step of the interpretation. Obviously information may be lost in this way but, as has been shown in the past, much confusion caused by an attempt to interpret the scattering near the Bragg peaks with specific structural properties of a cluster or molecular model is avoided. The inclusion of this part of diffuse scattering can be made after a complete interpretation of the change of the influence of the distribution function on diffraction in the wide-angle region.

4.2.5. Quantitative interpretation

4.2.5.1. Introduction

In these sections, quantitative interpretations of the elastic part of X-ray and neutron diffuse scattering are outlined. Although similar relations are valid for the magnetic scattering of neutrons, this particular topic is excluded. Obviously, all disorder phenomena are strongly temperature-dependent if thermal equilibrium is reached. Consequently, the interpretation of diffuse scattering should include a statistical thermodynamical treatment. Unfortunately, no quantitative theory for the interpretation of structural phenomena is so far available: all quantitative solutions introduce formal order parameters such as correlation functions or distributions of defects. At low temperatures (*i.e.* with a low concentration of defects) the distribution function plays the dominant role in diffuse scattering. With increasing temperature the number of defects increases with corresponding strong interactions between them. Therefore, correlations become increasingly important, and phase transformations of first or higher order may occur which need a separate theoretical treatment. In many cases, large fluctuations of structural properties occur that are closely related to the dynamical properties of the crystal. Theoretical approximations are possible, but their presentation is far beyond the scope of this article. Hence we restrict ourselves to formal parameters in the following.

Point defects or limited structural units, such as molecules, clusters of finite size *etc.*, may only be observed in diffraction if there is a sufficiently large number of defects. This statement is no longer true in high-resolution electron diffraction, where single defects may be observed either by diffraction or by optical imaging if their contrast is high enough. Hence electron microscopy and diffraction are valuable methods in the interpretation of disorder phenomena.

The arrangement of a finite assembly of structural defects is described by its structure and its three-dimensional (3D) distribution function. Structures with a strict 1D periodicity (chain-like structures) need a 2D distribution function, while for structures with a 2D periodicity (layers) a 1D distribution function is sufficient. Since the distribution function is the dominant factor in statistics with correlations between defects, we define the dimensionality of disorder as that of the corresponding distribution function. This definition is more effective in diffraction problems because the dimension of the disorder problem determines the dimension of the diffuse scattering: 1D diffuse streaks, 2D diffuse layers or a general 3D diffuse scattering.

Strictly speaking, completely random distributions cannot be realized, as shown in Section 4.2.3. They occur approximately if the following conditions are satisfied:

(1) The average volume of all defects including their surrounding strain fields NcV_d (where N is the number of unit cells, c is the concentration of defects and V_d is the volume of the defect with $V_d > V_c$, V_c being the volume of the unit cell) is small in comparison with the total volume NV_c of the crystal, or $V_c \gg cV_d$.

(2) Interactions between the defects are negligible. These conditions, however, are valid in very rare cases only, *i.e.* where small concentrations and vanishing strain fields are present. Remarkable exceptions from this rule are real point defects without interactions, such as isotope distribution (which can be studied by neutron diffraction) or the system AuAg at high temperature.

As already mentioned, disorder phenomena may be observed in thermal equilibrium. Two completely different cases have to be considered:

(1) The concentration of defects is given by the chemical composition, *i.e.* impurities in a closed system.

(2) The number of defects increases with temperature and also depends on pressure or other parameters, *i.e.* interstitials, voids, static displacements of atoms, stacking faults, dislocations *etc.*

In many cases, the defects do not occur in thermal equilibrium. Nevertheless, their diffuse scattering is temperature-dependent because of the anomalous thermal movements at the boundary of the defect. Hence the observation of a temperature-dependent behaviour of diffuse scattering cannot be taken as a definite criterion of thermal equilibrium without further careful interpretation.

Ordering of defects may take place in a very anisotropic manner. This is demonstrated by the huge number of examples of 1D disorder. As shown by Jagodzinski (1963), this type of disorder cannot occur in thermal equilibrium for the infinite crystal. This type of disorder is generally formed during crystal growth or mechanical deformation. Similar arguments may be applied to 2D disorder. This is a further reason why the Ising model can hardly ever be used to obtain interaction energies of structural defects. From these remarks it becomes clear that order parameters are formal parameters without strict thermodynamical meaning.

The following section is organized as follows: first we discuss the simple case of 1D disorder where reliable solutions of the diffraction problem are available. Intensity calculations for diffuse scattering from 2D disorder by chain-like structures follow. Finally, the 3D case is treated, where formal solutions of the diffraction problem have been tried and applied successfully to metallic systems to some extent. A short concluding section concerns the special phenomenon of orientational disorder.

4.2.5.2. Layered structures: one-dimensional disorder

As has been pointed out above, it is often useful to start the interpretation of diffuse scattering by checking the diffraction pattern with respect to the dimensionality of the disorder concerned. Since each disordered direction in the crystal demands a violation of the corresponding Laue condition, this question may easily be answered by looking at the diffuse scattering. Diffuse streaks in reciprocal space are due to a one-dimensional violation of the Laue conditions, and will be called one-dimensional disorder. This kind of order is typical for layer structures, but it is frequently observed in cases where several sequences of layers do not differ in the interactions of next-nearest neighbours. Typical examples are structures which may be described in terms of close packing, *e.g.* hexagonal and cubic close packing.

For a quantitative interpretation of diffuse streaks we need one-dimensional correlation functions, which may be determined uniquely if a single independent correlation function is active. According to equation (4.2.3.50), Fourier transformation yields the information required. In all other cases, a specific model has to be suggested for a full interpretation of diffuse streaks. It is worth noting that disorder parameters can be defined uniquely

only if the diffraction pattern allows for a differentiation between long-range and short-range order. This question can be answered at least partly by studying the line width of sharp reflections at very good resolution. Since integrated intensities of sharp reflections have to be separated from the diffuse scattering, this question is of outstanding importance in most cases. Inclusion of diffuse parts in the diffraction pattern during intensity integration of sharp reflections may lead to serious errors in the interpretation of the average structure.

The existence of diffuse streaks in more than one direction of reciprocal space means that the diffraction problem is no longer one-dimensional. Sometimes the problem may be treated independently if the streaks are sharp and no interference effects are observed in the diffraction pattern in areas where the diffuse streaks do overlap. In all other cases, there are correlations between the various directions of one-dimensional disorder which can be determined with the aid of a model covering more than one of the pertinent directions of disorder.

Before starting the discussion of the quantitative solution of the one-dimensional problem, some remarks should be made about the usefulness of quantitative disorder parameters. It is well known from statistical thermodynamics that a one-dimensional system cannot show long-range order above $T = 0$ K. Obviously, this statement is in contradiction with many experimental observations of long-range order even in layer structures. The reason for this behaviour is given by the following arguments, which are valid for any structure. Let us assume a structure with strong interactions at least in two directions. From the theoretical treatment of the two-dimensional Ising model it is known that such a system shows long-range order below a critical temperature T_c. This statement is true even if the layer is finite, although the strict thermodynamic behaviour is not really critical in the thermodynamical sense. A three-dimensional crystal can be constructed by adding layer after layer. Since each layer has a typical two-dimensional free energy, the full statistics of the three-dimensional crystal may be calculated by introducing a specific free energy for the various stackings of layers. Obviously, this additional energy also has to include terms describing potential and entropic energies. They may be formally developed into contributions of next, next-but-one *etc.* nearest neighbours. The contribution to entropy must include configurational and vibrational parts, which are strongly coupled. As long as the layers are finite, there is a finite probability of a fault in the stacking sequence of layers which approaches zero with increasing extension of the layers. Consequently, the free energy of a change in the favourite stacking sequence becomes infinite quadratically with the size of the layer. Therefore, the crystal should be either completely ordered or disordered; the latter case can only be realized if the free energies of one or more stacking sequences are exactly equal (this is very rare but is possible over a small temperature range during phase transformations). An additional positive entropy associated with a deviation from the periodic stacking sequence may lead to a kind of competition between entropy and potential energy, in such a way that periodic sequences of faults result. This situation occurs in the transition range of two structures differing only in their stacking sequence. On the other hand, one must assume that defects in the stacking sequence may occur if the size of the layers is small. This situation occurs during crystal growth, but one should remember that the number of stacking defects should decrease with increasing size of the growing crystal. Apparently, this rearrangement of layers may be suppressed as a consequence of relaxation effects. The growth process itself may influence the propagation of stacking defects and, consequently, the determination of stacking-fault probabilities, with the aim of interpreting the chemical bonding, seems to be irrelevant in most cases.

The quantitative solution of the diffraction problem of one-dimensional disorder follows a method similar to the Ising model. As long as next-nearest neighbours alone are considered, the solution is very simple if only two possibilities for the structure factors are to be taken into account. Introducing the probability of equal pairs 1 and 2, α, one arrives at the known solution for the *a priori* probability p_μ and *a posteriori* probabilities $p_{\mu\mu'}(\mathbf{m})$, respectively. In the one-dimensional Ising model with two spins and the interaction energies $(U \mp \Delta U)/k_B T$, defining the pair probability $[\alpha = p_{11}(\mathbf{1})]$ as

$$\alpha = \frac{\exp\{\pm \Delta U/k_B T\}}{[\exp\{+\Delta U/k_B T\} + \exp\{-\Delta U/k_B T\}]},$$

the full symmetry is $p_1 = p_2 = \frac{1}{2}$ and $p_{11}(\mathbf{m}) = p_{22}(\mathbf{m})$.

Consequently

$$p_{12}(\mathbf{m}) = p_{22}(\mathbf{m}) = 1 - p_{11}(\mathbf{m}).$$

The scattered intensity is given by

$$I(\mathbf{H}) = L(h,k) \sum_{\mathbf{m}} \langle FF_m^+ \rangle (N - |m|) \exp\{-2\pi i m l\}, \qquad (4.2.5.1)$$

where $\mathbf{m} = m\mathbf{c}$, N is the number of unit cells in the $\mathbf{c}$ direction and $\langle FF^+ \rangle$ depends on λ_1, λ_2, which are the eigenvalues of the matrix

$$\begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}.$$

From the characteristic equation

$$\lambda^2 - 2\alpha\lambda - 1 + 2\alpha = 0 \qquad (4.2.5.2)$$

one has

$$\lambda_1 = 1; \quad \lambda_2 = 2\alpha - 1. \qquad (4.2.5.2a)$$

λ_1 describes a sharp Bragg reflection (from the average structure) which need not be calculated. Its intensity is simply proportional to $\langle F(\mathbf{H}) \rangle$. The second characteristic value yields a diffuse reflection in the same position if the sign is positive ($\alpha > 0.5$) and in a position displaced by $\frac{1}{2}$ in reciprocal space if the sign is negative ($\alpha < 0.5$). Because of the symmetry conditions only $p_{11}(\mathbf{m})$ is needed; it may be determined with the aid of the boundary conditions

$$p_{11}(\mathbf{0}) = 1, \quad p_{11}(\mathbf{1}) = \alpha,$$

and the general relation

$$p_{\mu\mu'}(\mathbf{m}) = c'_{\mu\mu'} \lambda_1^m + c''_{\mu\mu'} \lambda_2^m.$$

The final solution of our problem yields simply

$$p_{11}(\mathbf{m}) = \tfrac{1}{2} + \tfrac{1}{2}\lambda_2^m = p_{22}(\mathbf{m}),$$
$$p_{12}(\mathbf{m}) = \tfrac{1}{2} - \tfrac{1}{2}\lambda_2^m = p_{21}(\mathbf{m}).$$

The calculation of the scattered intensity is now performed with the general formula

$$I(\mathbf{H}) = L(h,k) \sum_m \sum_{\mu,\mu'} p_\mu p_{\mu\mu'}(\mathbf{m}) F_\mu F_{\mu'}^+ (N - |m|) \exp\{-2\pi i m l\}. \qquad (4.2.5.3)$$

Evaluation of this expression yields

$$I(\mathbf{H}) = L(h,k) \sum_m (N - |m|) \exp\{-2\pi i m l\} \times [|\tfrac{1}{2}(F_1 + F_2)|^2 \lambda_1^m + |\tfrac{1}{2}(F_1 - F_2)|^2 \lambda_2^m]. \qquad (4.2.5.4)$$

Since the characteristic solutions of the problem are real,

$$I(\mathbf{H}) = L(\mathbf{G})|(F_1 + F_2)/2|^2 + L(h,k)|(F_1 - F_2)/2|^2 \times \frac{1 - |\lambda_2|^2}{1 - 2\lambda_2 \cos 2\pi l + |\lambda_2|^2}. \quad (4.2.5.5)$$

The particle-size effect has been neglected in (4.2.5.5). This result confirms the fact mentioned above that the sharp Bragg peaks are determined by the averaged structure factor and the diffuse ones by its mean-square deviation.

For the following reason there are no examples of quantitative applications: two different structures generally have different lattice constants, so the original assumption of an undisturbed lattice geometry is not valid. The only case known to the authors is the typical lamellar structure of plagioclases reported by Jagodzinski & Korekawa (1965). The authors interpret the well known 'Schiller effect' as a consequence of optical diffraction. Hence, the size of the lamellae is of the order of 2000 Å. This long-period superstructure cannot be explained in terms of next-nearest-neighbour interactions. In principle, however, the diffraction effects are similar: instead of the diffuse peak as described by the second term in equation (4.2.5.5), satellites of first and second order *etc.* accompanying the Bragg peaks are observed. The study of this phenomenon (Korekawa & Jagodzinski, 1967) has in the meantime found a quantitative interpretation (Burandt *et al.*, 1992; Kalning *et al.*, 1997).

Obviously, the symmetry relation used in the formulae discussed above is only valid if the structures described by the F_μ are related by symmetries such as translations, rotations or combinations of the two. The type of symmetry has an important influence on the diffraction pattern.

(1) *Translation parallel to the ordered layers*

If the translation vector between the two layers in question is such that $2\Delta\mathbf{r}$ is a translation vector parallel to the layer, there are two relevant structure factors:

$$F_1, F_2 = F_1 \exp\{2\pi i \mathbf{H} \cdot \Delta\mathbf{r}\}.$$

$\mathbf{H} \cdot \Delta\mathbf{r}$ may be either an integer or an integer + $\frac{1}{2}$. Since any integer may be neglected because of the translation symmetry parallel to the layer, we have $F_1 = F_2$ in the former case and $F_1 = -F_2$ in the latter. As a consequence, either the sharp reflections given in equation (4.2.5.4) vanish, or the same is true for the diffuse ones. Hence the reciprocal lattice may be described in terms of two kinds of lattice rows, sharp and diffuse, parallel to the reciprocal coordinate l.

Disorder of this type is observed very frequently. One of the first examples was wollastonite, $CaSiO_3$, studied by Jefferey (1953). Here the reflections with $k = 2n$ are sharp Bragg peaks without any diffuse scattering. Diffuse streaks parallel to (h00), however, are detected for $k = 2n + 1$. In the light of the preceding discussion, the translation vector is $\frac{1}{2}\mathbf{b}$ and the plane of ordered direction (the plane of intergrowth of the two domains) is (100). Hence the displacement is parallel to this plane. Since the intensity of the diffuse lines does not vary according to the structure factor involved, the disorder cannot be random. The maxima observed are approximately in the position of a superstructure generated by large domains without faults in the stacking sequence mutually displaced by $\frac{1}{2}\mathbf{b}$ (antiphase domains). This complicated ordering behaviour is typical for 1D order and may easily be explained by the above-mentioned fact that an infinitely extended interface between two domains causes an infinite unfavourable energy (Jagodzinski, 1964*b*, p. 188). Hence a growing crystal should become increasingly ordered. This consideration explains why the agreement between a 1D disorder theory and experiment is often so poor.

Examples where more than one single displacement vector occur are common. If these are symmetrically equivalent all symmetries have to be considered. The most important cases of displacements differing only by translation are the well known close-packed structures (see below). A very instructive example is the mineral maucherite (approximately Ni_4As_3). According to Jagodzinski & Laves (1947), the structure has the following disorder parameters: interface (001), displacement vectors [000], [$\frac{1}{2}$00], [0$\frac{1}{2}$0], [$\frac{1}{2}\frac{1}{2}$0]. From equation (4.2.5.5) we obtain

$$\langle F(\mathbf{H})\rangle = [1 + \exp\{\pi i h\} + \exp\{\pi i k\} + \exp\{\pi i (h + k)\}]/4.$$

Hence there are sharp reflections for h, k = even and diffuse ones otherwise. Further conclusions may be drawn from the average structure.

(2) *Translation perpendicular to the ordered layers*

If the translation is $\mathbf{c}/2$ the structure factors are

$$F_2 = F_1 \exp\{2\pi i l\},$$
$$F_1 = F_2 \text{ for } l = \text{even},$$
$$F_1 = -F_2 \text{ for } l = \text{odd}.$$

There are sharp ($l = 2n$) *and* diffuse ($l = 2n + 1$) reflections on all reciprocal-lattice rows discussed above.

Since the sharp and diffuse reflections occur on the same reciprocal line, the behaviour is completely different compared with the preceding case. In general, a component of any displacement vector perpendicular to the interface gives rise to a change in chemical composition, as shown in the next example: in a binary system consisting of A and B atoms with a tendency towards an alternating arrangement of A and B layers, any fault in the sequence $BABAB|BABAB|B$ increases the number of B atoms (or A atoms). In general, such kinds of defects will lead to an interface with a different lattice constant, at least in the direction perpendicular to the interface. Consequently, the exact displacement vectors of $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ are rare. Since ordered structures should be realized in the 1D case, incommensurate superstructures will occur; these are very abundant during ordering processes. An interesting example has been reported and interpreted by Cowley (1976*a*,*b*), where the displacement vector has a translational period of $\frac{1}{4}$ perpendicular to the plane of intergrowth. Reflections 00l and 22l with l = 4n are sharp, all remaining reflections more or less diffuse. Since the maxima ($\bar{1}11$), ($\bar{1}33$) show a systematically different behaviour, there is also a displacement component parallel to the plane of intergrowth in question. A semiquantitative interpretation was given in his papers.

(3) *Rotations*

Layers related by a twofold rotation parallel to **c** may easily be discussed by simply considering their structure factors. Since the layers do not obey the twofold symmetry, their structure factors are generally different; unless they become equal accidentally there are sharp and diffuse reflections according to the values of $\langle F\rangle$ and ΔF, respectively. Obviously, $F_1 = F_2$ is valid only if $h = k = 0$; consequently there is just one reciprocal-lattice row free from diffuse scattering.

(4) *Asymmetric case*

In the asymmetric case, the symmetry conditions used above are no longer valid:

$$p_1 \neq p_2, \quad p_{12} \neq p_{21}, \quad p_{11} \neq p_{22}.$$

However, there is one condition which may be derived from the invariance of the numbers of pairs in the relevant and its opposite direction:

$$p_\mu p_{\mu\mu'}(\mathbf{m}) = p_{\mu'} p_{\mu'\mu}(-\mathbf{m}).$$

This equation does not require that $p_{\mu\mu'}(\mathbf{m})$ should be symmetric in $\mathbf{m}$. The calculation of characteristic values yields

$$\lambda_1 = 1, \qquad \lambda_2 = (\alpha_1 + \alpha_2) - 1. \quad (4.2.5.6)$$

The *a priori* probabilities are now different from $\frac{1}{2}$ and may be calculated by considering $p_{\mu\mu'}(\mathbf{m}) \rightarrow p_{\mu'}(\mathbf{m} \rightarrow \infty)$:

$$p_1 = \alpha_1/(\alpha_1 + \alpha_2); \quad p_2 = \alpha_2/(\alpha_1 + \alpha_2).$$

The intensity is given by an expression very similar to (4.2.5.5):

$$\begin{aligned} I(\mathbf{H}) = {} & L(\mathbf{G})|[\alpha_1/(\alpha_1 + \alpha_2)]F_1 + [\alpha_2/(\alpha_1 + \alpha_2)]F_2|^2 \\ & + L(h,k)|[\alpha_1/(\alpha_1 + \alpha_2)]F_1 - [\alpha_2/(\alpha_1 + \alpha_2)]F_2|^2 \\ & \times (1 - |\lambda_2|^2)/(1 - 2\lambda_2 \cos 2\pi l + |\lambda_2|^2). \end{aligned} \tag{4.2.5.7}$$

Again, there are sharp Bragg reflections and diffuse ones in the same positions or in a displaced position depending on the sign of λ_2.

From a discussion of the next-nearest-neighbour Ising model, one may conclude that a detailed study of the qualitative behaviour of sharp and diffuse reflections may give additional information on the symmetry of the layers involved.

In the case of translations between neighbouring layers not fulfilling the condition $\mathbf{h} \cdot \mathbf{r}$ = integer, where $\mathbf{r}$ is parallel to the layer, more than two structure factors have to be taken into account. If $n\mathbf{h} \cdot \mathbf{r}$ = integer, where n is the smallest integer fulfilling the said condition, n different structure factors have to be considered. The characteristic equation has formally to be derived with the aid of an $n \times n$ matrix containing internal symmetries, which may be avoided by adding the phase factors $\varepsilon = \exp\{2\pi i \mathbf{H} \cdot \mathbf{r}/n\}$, $\varepsilon^+ = \exp\{-2\pi i \mathbf{H} \cdot \mathbf{r}/n\}$ to the probability of pairs. The procedure is allowed if the displacements $\mathbf{r}$ and $-\mathbf{r}$ are admitted only for neighbouring layers. The matrix yielding the characteristic values may then be reduced to

$$\begin{pmatrix} \alpha_1\varepsilon & (1-\alpha_1)\varepsilon^+ \\ (1-\alpha_2)\varepsilon & \alpha_2\varepsilon^+ \end{pmatrix}$$

and yields the characteristic equation

$$\lambda^2 - \lambda(\alpha_1\varepsilon + \alpha_2\varepsilon^+) - 1 + \alpha_1 + \alpha_2 = 0. \tag{4.2.5.8}$$

Equation (4.2.5.8) gives sharp Bragg reflections for $\mathbf{H} \cdot \mathbf{r}/n$ = integer; the remaining diffuse reflections are displaced according to the phase of the complex characteristic value. Equation (4.2.5.8) has been used in many cases. Qualitative examples are the mixed-layer structures published by Hendricks & Teller (1942). An example of a four-layer-type structure is treated by Dubernat & Pezerat (1974). A first quantitative treatment with good agreement between theory and experimental powder diffraction data has been given by Dorner & Jagodzinski (1972) for the binary system TiO_2–SnO_2. In the range of the so-called spinodal decomposition, the chemical compositions of the two domains and the average lengths of the two types of domains could be determined. Another quantitative application was reported by Jagodzinski & Hellner (1956) for the transformation of $RhSn_2$ into a very complicated mixed-layer type. A good agreement of measured and calculated diffuse scattering (with asymmetric line profiles and displacement of maxima) could be found over a wide angular range of single-crystal diffraction.

4.2.5.2.1. Stacking disorder in close-packed structures

From an historical point of view, stacking disorder in close-packed systems is most important. The three relevant positions of ordered layers are represented by the atomic coordinates $|0,0|$, $|\frac{1}{3},\frac{2}{3}|$, $|\frac{2}{3},\frac{1}{3}|$ in the hexagonal setting of the unit cell, or simply by the figures 1, 2, 3 in the same sequence. Structure factors F_1, F_2, F_3 refer to the corresponding positions of the same layer:

$$F_2 = F_1 \exp\{2\pi i(h-k)/3\},$$
$$F_3 = F_1 \exp\{-2\pi i(h-k)/3\},$$

hence

$$F_1 = F_2 = F_3 \quad \text{if } h - k \equiv 0 \pmod 3.$$

According to the above discussion, the indices 1, 2, 3 define the reciprocal-lattice rows exhibiting sharp reflections only, as long as the distances between the layers are exactly equal. The symmetry conditions caused by the translation are normally

$$p_1 = p_2 = p_3, \qquad p_{11} = p_{22} = p_{33},$$
$$p_{12} = p_{23} = p_{31}, \quad p_{13} = p_{21} = p_{32}.$$

For the case of close packing of spheres and some other problems any configuration of $\mathbf{m}$ layers determining the *a posteriori* probability $p_{\mu\mu'}(\mathbf{m})$, $\mu = \mu'$, has a symmetrical counterpart where μ is replaced by $\mu' + 1$ (if $\mu' = 3$, $\mu' + 1 = 1$).

In this particular case, $p_{12}(\mathbf{m}) = p_{13}(\mathbf{m})$ and equivalent relations generated by translation.

Nearest-neighbour interactions do not lead to an ordered structure if the principle of close packing is obeyed (no pairs are in equal positions) (Hendricks & Teller, 1942; Wilson, 1942). Extension of the interactions to next-but-one or more neighbours may be carried out by introducing the method of matrix multiplication developed by Kakinoki & Komura (1954, 1965) or the method of overlapping clusters (Jagodzinski, 1954). The latter procedure is outlined in the case of interactions between four layers. A given set of three layers may occur in the following 12 combinations:

$$123, 231, 312; \quad 132, 213, 321;$$
$$121, 232, 313; \quad 131, 212, 323.$$

Since three of them are equivalent by translation, only four representatives have to be introduced:

$$123; \quad 132; \quad 121; \quad 131.$$

In the following, the new indices 1, 2, 3, 4 are used for these four representatives for the sake of simplicity.

In order to construct the statistics layer by layer, the next layer must belong to a triplet starting with the same two symbols with which the preceding one ended, *e.g.* 123 can only be followed by 231 or 232. In a similar way, 132 can only be followed by 321 or 323. Since both cases are symmetrically equivalent, the probabilities α_1 and $1 - \alpha_1$ are introduced. In a similar way, 121 may be followed by 212 or 213 *etc.* For these two groups the probabilities α_2 and $1 - \alpha_2$ are defined. The different translations of groups are considered by introducing the phase factors as described above. Hence the matrix for the characteristic equation may be set up as follows. As a representative cluster of each group the one having the number 1 at the centre is chosen, *e.g.* 312 is the representative for the group 123, 231, 312; in a similar way 213, 212 and 313 are the remaining representatives. Since this arrangement of three layers is equivalent by translation, it may be assumed that the structure of the central layer is not influenced by the statistics to a first approximation. The same arguments hold for the remaining three groups. On the other hand, the groups 312 and 213 are equivalent by rotation only. Consequently, their structure factors may differ if the influence of the two neighbours has to be taken into account. A different situation exists for the groups 212 and 313, which are correlated by a centre of symmetry, which causes different corresponding structure factors. It should be pointed out, however, that the structure factor is invariant as long as there is no influence of neighbouring layers on the structure of the central layer. The latter is often observed in close-packed metal structures or in compounds like

ZnS, SiC and others. For the calculation of intensities $p_\mu p_{\mu\mu'}$ and $F_\mu F^+_{\mu'}$ are needed.

According to the following scheme of sequences, any sequence of pairs is correlated with the same phase factor for FF^+ due to translation if both members of the pair belong to the same group. Consequently, the phase factor may be attached to the sequence probability such that FF^+ remains unchanged and the group may be treated as a single element in the statistics. In this way, the reduced matrix for the solution of the characteristic equation is given by

F_μ	$F^+_{\mu'}$ (1) 312, 123(ε^+), 231(ε)	(2) 212, 323(ε^+), 131(ε)	(3) 213, 321(ε^+), 132(ε)	(4) 313, 121(ε^+), 232(ε)
(1) 312, 123(ε), 231(ε^+)	$\alpha_1\varepsilon^+$	0	0	$(1-\alpha_1)\varepsilon^+$
(2) 212, 323(ε), 131(ε^+)	$(1-\alpha_2)\varepsilon^+$	0	0	$\alpha_2\varepsilon^+$
(3) 213, 321(ε), 132(ε^+)	0	$(1-\alpha_2)\varepsilon$	$\alpha_1\varepsilon$	0
(4) 313, 121(ε), 232(ε^+)	0	$\alpha_2\varepsilon$	$(1-\alpha_2)\varepsilon$	0

It should be pointed out that clusters 1 and 3 describe a cubic arrangement of three layers in the case of simple close packing, while clusters 2 and 4 represent the hexagonal close packing. There may be a small change in the lattice constant c perpendicular to the layers. Additional phase factors then have to be introduced in the matrix for the characteristic equation and a recalculation of the constants is necessary. As a consequence, the reciprocal-lattice rows $(h-k)\equiv 0\,(\mathrm{mod}\ 3)$ become diffuse if $l\neq 0$ and the diffuseness increases with l. Similar behaviour results for the remaining reciprocal-lattice rows.

There are three solutions of the diffraction problem:

(1) If $h - k = 0$ (mod 3), $\varepsilon = +1$, there are two quadratic equations:

$$\begin{aligned}\lambda^2-(\alpha_1+\alpha_2)\lambda-1+\alpha_1+\alpha_2&=0,\\ \lambda^2-(\alpha_1-\alpha_2)\lambda+1-\alpha_1-\alpha_2&=0,\end{aligned}\qquad(4.2.5.9)$$

with solutions

$$\lambda_1=1,\qquad \lambda_2=\alpha_1+\alpha_2-1,$$
$$\lambda_{3/4}=\frac{\alpha_1-\alpha_2}{2}\pm\left[\frac{(\alpha_1-\alpha_2)^2}{4}-1+\alpha_1-\alpha_2\right]^{1/2}.\qquad(4.2.5.10)$$

λ_1 and λ_2 are identical with the solution of the asymmetric case of two kinds of layers [*cf.* equation (4.2.5.6)]. They yield sharp reflections for l = integer and diffuse ones in a position determined by the sequence probabilities α_1 and α_2 (with a position of either l = integer or $l = \frac{1}{2}$ + integer, respectively). The remaining two characteristic values may be given in the form $\lambda = |\lambda|\exp\{2\pi i\varphi\}$, where φ determines the position of the reflection. If the structure factors of the layers are independent of the cluster, λ_2, λ_3 and λ_4 become irrelevant because of the new identity of the F's and there is no diffuse scattering. Weak diffuse intensities on the lattice rows $h - k = 0$ (mod 3) may be explained in terms of this influence.

(2) The remaining two solutions for $\varepsilon = \exp\{\pm 2\pi i(h-k)/3\}$ are equivalent and result in the same characteristic values. They have been discussed explicitly in the literature; the reader is referred to the papers of Jagodzinski (1949*a*,*b*,*c*, 1954).

In order to calculate the intensities, one has to reconsider the symmetry of the clusters, which is different to the symmetry of the layers. Fortunately, a threefold rotation axis is invariant against the translations, but this is not true for the remaining symmetry operations in the layer if there are any more. Since we have two pairs of inequivalent clusters, namely 312, 213 and 212, 313, there are only two different *a priori* probabilities $p_1 = p_3$ and $p_2 = p_4 = \frac{1}{2}(1 - 2p_1)$.

The symmetry conditions of the new clusters may be determined using 'probability trees' as described by Wilson (1942) and Jagodzinski (1949*b*). For example: $p_{11} = p_{33}$, $p_{22} = p_{44}$, $p_{13} = p_{31}$, $p_{24} = p_{42}$ *etc*.

The final solution of the diffraction problem results in the following general intensity formula:

$$\begin{aligned}I(\mathbf{H}) = L(h,k)N\sum_\nu\{&A_\nu(\mathbf{H})(1-|\lambda_\nu|^2)\\ &\times[1-2|\lambda_\nu|\cos 2\pi(l-\varphi_\nu)+|\lambda_\nu|^2]^{-1}\\ &-2B_\nu(\mathbf{H})|\lambda_\nu|\sin 2\pi(l-\varphi_\nu)\\ &\times[1-2|\lambda_\nu|\cos 2\pi(l-\varphi_\nu)+|\lambda_\nu|^2]^{-1}\}.\end{aligned}\qquad(4.2.5.11)$$

Here A_ν and B_ν represent the real and imaginary part of the constants to be calculated with the aid of the boundary conditions of the problem. The first term in equation (4.2.5.11) determines the symmetrical part of a diffuse reflection with respect to the maximum and is completely responsible for the integrated intensity. The second term causes an antisymmetrical contribution to intensity profiles but does not influence the integrated intensities. These general relations enable a semiquantitative interpretation of the sharp and diffuse scattering in any case, without performing the time-consuming calculations of the constants, which may only be done in more complicated disorder problems with the aid of a computer program evaluating the boundary conditions of the problem.

This can be carried out with the aid of the characteristic values and a linear system of equations (Jagodzinski, 1949*a*,*b*,*c*), or with the aid of matrix formalism (Kakinoki & Komura, 1954; Takaki & Sakurai, 1976). As long as only the line profiles and positions of the reflections are required, these quantities may be determined experimentally and fitted to characteristic values of a matrix. The size of this matrix is given by the number of sharp and diffuse maxima observed, while $|\lambda_\nu|$ and $\exp\{2\pi i\varphi_\nu\}$ may be found by evaluating the line width and the position of diffuse reflections. Once this matrix has been found, a semiquantitative model of the disorder problem can be given. If a system of sharp reflections is available, the averaged structure can be solved as described in Section 4.2.3.2. The determination of the constants of the diffraction problem is greatly facilitated by considering the intensity modulation of diffuse scattering, which enables a phase determination of structure factors to be made under certain conditions.

The theory of close-packed structures with three equivalent translation vectors has been applied very frequently, even to

systems that do not obey the principle of close-packing. The first quantitative explanation was published by Halla *et al.* (1953). It was shown there that single crystals of $C_{18}H_{24}$ from the same synthesis may have a completely different degree of order. This was true even within the same crystal. Similar results were found for C, Si, CdI_2, CdS_2, mica and many other compounds. Quantitative treatments are less abundant [*e.g.* CdI_2 (Martorana *et al.*, 1986); MX_3 structures (Conradi & Müller, 1986)]. Special attention has been paid to the quantitative study of polytypic phase transformations in order to gain information about the thermodynamical stability or the mechanism of layer displacements, *e.g.* Co (Edwards & Lipson, 1942; Frey & Boysen, 1981), SiC (Jagodzinski, 1972; Pandey *et al.*, 1980*a*,*b*,*c*), ZnS (Müller, 1952; Mardix & Steinberger, 1970; Frey *et al.*, 1986) and others.

Certain laws may be derived for the reduced integrated intensities of diffuse reflections. 'Reduction' in this context means a division of the diffuse scattering along l by the structure factor, or the difference structure factor if $\langle F\rangle \neq 0$. This procedure is valuable if the number of stacking faults rather than a complete solution of the diffraction problem is required.

The discussion given above has been made under the assumption that the full symmetry of the layers is maintained in the statistics. Obviously, this is not necessarily true if external lower symmetries influence the disorder. An important example is the generation of stacking faults during plastic deformation. Problems of this kind need a complete reconsideration of symmetries. Furthermore, it should be pointed out that a treatment with the aid of an extended Ising model as described above is irrelevant in most cases. Simplified procedures describing the diffuse scattering of intrinsic, extrinsic, twin stacking faults and others have been described in the literature. Since their influence on *structure determination* can generally be neglected, the reader is referred to the literature for additional information.

A different approach to the analysis of planar disorder in close-packed structures is given by Varn *et al.* (2002). With simplifying assumptions, *e.g.* that spacings between identical defect-free layers are independent of local stacking arrangements, average correlations between them are extracted from experimental diffractograms *via* Fourier analysis. Spatial patterns of the layers are then constructed by a so-called epsilon machine, which reproduces the correlations with minimal states of the stacking process. The basic statistical description of the ensemble of spatial patterns produces the stacking distribution. With this technique, stacking sequences are generated which are compared with the correlation factors and the diffractograms. The authors report an improved matching of calculated and experimental data for ZnS.

4.2.5.3. Chain-like structures: two-dimensional disorder

In this section, disorder phenomena that are related to chain-like structural elements in crystals are considered. This topic includes the so-called '1D crystals', where translational symmetry (in direct space) exists in one direction only – crystals in which highly anisotropic binding forces are responsible for chain-like atomic groups, *e.g.* compounds that exhibit a well ordered 3D framework structure with tunnels in a unique direction in which atoms, ions or molecules are embedded. Examples are compounds with platinum, iodine or mercury chains, urea inclusion compounds with columnar structures (organic or inorganic), 1D ionic conductors, polymers *etc.* Diffuse-scattering studies of 1D conductors have been carried out in connection with investigations of stability/instability problems, incommensurate structures, phase transitions, dynamic precursor effects *etc.* These areas are not treated here. For general reading about diffuse scattering in connection with these topics see, *e.g.*, Comes & Shirane (1979) and references therein. Also excluded are specific problems related to polymers or liquid crystals (mesophases) (see Chapter 4.4) and magnetic structures with chain-like spin arrangements.

Trivial diffuse scattering occurs as 1D Bragg scattering (diffuse layers) by internally ordered chains. Diffuse phenomena in reciprocal space are due to 'longitudinal' disordering within the chains (along the unique direction) as well as to 'transverse' correlations between different chains over a restricted volume. Only static aspects are considered; diffuse scattering resulting from collective excitations or diffusion-like phenomena which are of inelastic or quasielastic origin are not treated here.

4.2.5.3.1. Randomly distributed collinear chains

As found in any elementary textbook of diffraction, the simplest result of scattering by a chain with period c,

$$l(\mathbf{r}) = l(z) = \sum_{n_3} \delta(z - n_3 c), \tag{4.2.5.12}$$

is described by one of the Laue equations:

$$G(L) = |L(L)|^2 = \sin^2 \pi NL / \sin^2 \pi L, \tag{4.2.5.13}$$

which gives broadened profiles for small N. In the context of phase transitions the Ornstein–Zernike correlation function is frequently used, *i.e.* (4.2.5.13) is replaced by a Lorentzian:

$$1/[\xi^2 + 4\pi^2 (L - l)^2], \tag{4.2.5.14}$$

where ξ denotes the correlation length.

In the limiting case $N \rightarrow \infty$, (4.2.5.13) becomes

$$\sum_l \delta(L - l). \tag{4.2.5.15}$$

The scattering by a real chain $a(\mathbf{r})$ consisting of molecules with structure factor F_M is therefore determined by

$$F_M(\mathbf{H}) = \sum_j f_j \exp\{2\pi i(Hx_j + Ky_j + Lz_j)\}. \tag{4.2.5.16}$$

The Patterson function is

$$\begin{aligned} P(\mathbf{r}) = {} & (1/c) \int\int |F_0(H, K)|^2 \cos 2\pi(Hx + Ky)\, \mathrm{d}H\, \mathrm{d}K \\ & + (2/c) \sum_l \int\int |F_l|^2 \exp\{2\pi i(Hx + Ky)\} \\ & \times \exp\{-2\pi i l z\}\, \mathrm{d}H\, \mathrm{d}K, \end{aligned} \tag{4.2.5.17}$$

where the index l denotes the only relevant position $L = l$ (the subscript M is omitted).

The intensity is concentrated in diffuse layers perpendicular to $\mathbf{c}^*$ from which the structural information may be extracted. Projections are:

$$\int a(\mathbf{r})\, \mathrm{d}z = \int\int F_0(H, K) \exp\{2\pi i(Hx + Ky)\}\, \mathrm{d}H\, \mathrm{d}K, \tag{4.2.5.18}$$

$$\int\int a(\mathbf{r})\, \mathrm{d}x\, \mathrm{d}y = (2/c) \sum_l F_l(00l) \exp\{-2\pi i l z\}. \tag{4.2.5.19}$$

Obviously the z parameters can be determined by scanning along a meridian $(00L)$ through the diffuse sheets (*e.g.* by a diffractometer recording). Owing to intersection of the Ewald sphere with the set of planes, the meridian cannot be recorded on *one* photograph; successive equi-inclination photographs are necessary. Only in the case of large c spacings is the meridian well approximated in one photograph.

There are many examples where a tendency to cylindrical symmetry exists: chains with p-fold rotational or screw symmetry around the preferred direction or assemblies of chains (or domains) with a statistical orientational distribution around the texture axis. In this context, it should be mentioned that symmetry operations with rotational parts belonging to the 1D rod groups actually occur, *i.e.* not only $p = 2, 3, 4, 6$.

In all these cases a treatment in the frame of cylindrical coordinates is advantageous (see, *e.g.*, Vainshtein, 1966):

Direct space	Reciprocal space
$x = r\cos\psi$	$H = H_r\cos\Psi$
$y = r\sin\psi$	$K = H_r\sin\Psi$
$z = z$	$L = L$

$$a(r,\psi,z) = \int\int\int F(\mathbf{H})\exp\{-2\pi i[H_r r\cos(\psi-\Psi)+Lz]\} \times H_r\,\mathrm{d}H_r\,\mathrm{d}\Psi\,\mathrm{d}L, \quad (4.2.5.20)$$

$$F(\mathbf{H}) = \int\int\int a(r,\psi,z)\exp\{2\pi i[H_r r\cos(\psi-\Psi)+Lz]\} \times r\,\mathrm{d}r\,\mathrm{d}\psi\,\mathrm{d}z. \quad (4.2.5.21)$$

The integrals may be evaluated by the use of Bessel functions:

$$J_n(u) = \tfrac{1}{2}\pi i^n\int\exp\{i(u\cos\varphi + n\varphi)\}\,\mathrm{d}\varphi$$

($u = 2\pi r H_r$; $\varphi = \psi - \Psi$).

The 2D problem $a = a(r,\psi)$ is treated first; an extension to the general case $a(r,\psi,z)$ is easily made afterwards.

Along the theory of Fourier series one has

$$a(r,\psi) = \sum_n a_n(r)\exp\{in\psi\},$$
$$a_n(r) = \frac{1}{2\pi}\int a(r,\psi)\exp\{-in\psi\}\,\mathrm{d}\psi, \quad (4.2.5.22)$$

or with

$$\alpha_n = \frac{1}{2\pi}\int a(r,\psi)\cos(n\psi)\,\mathrm{d}\psi,$$
$$\beta_n = \frac{1}{2\pi}\int a(r,\psi)\sin(n\psi)\,\mathrm{d}\psi,$$
$$a_n(r) = |a_n(r)|\exp\{-i\psi_n(r)\},$$
$$|a_n(r)| = \sqrt{a_n^2+\beta_n^2},$$
$$\psi_n(r) = \arctan\,\beta_n/\alpha_n.$$

If contributions to anomalous scattering are neglected $a(r,\psi)$ is a real function:

$$a(r,\psi) = \sum_n |a_n(r)|\cos[n\psi - \psi_n(r)]. \quad (4.2.5.23)$$

Analogously, one has

$$F(H_r,\Psi) = \sum_n |F_n(H_r)|\exp\{in\Psi\}. \quad (4.2.5.24)$$

$F(H_r,\Psi)$ is a complex function; $F_n(H_r)$ are the Fourier coefficients that are to be evaluated from the $a_n(r)$:

$$F_n(H_r) = \frac{1}{2\pi}\int F(H_r,\Psi)\exp\{-in\Psi\}\,\mathrm{d}\Psi$$
$$= \exp\{in\pi/2\}\int a_n(r)J_n(2\pi r H_r)2\pi r\,\mathrm{d}r,$$
$$F(H_r,\Psi) = \sum_n \exp\{in[\Psi+(\pi/2)]\} \times \int a_n(r)J_n(2\pi r H_r)2\pi r\,\mathrm{d}r, \quad (4.2.5.25)$$

$$a(r,\psi) = \sum_n \exp\{in[\Psi-(\pi/2)]\} \times \int F_n(H_r)J_n(2\pi r H_r)\,2\pi H_r\,\mathrm{d}H_r. \quad (4.2.5.26)$$

The formulae may be used for calculation of the diffuse intensity distribution within a diffuse sheet, in particular when the chain molecule is projected along the unique axis [*cf.* equation (4.2.5.18)].

Special cases are:

(1) *Complete cylinder symmetry*

$$F(H_r) = 2\pi\int a(r)J_0(2\pi r H_r)r\,\mathrm{d}r \quad (4.2.5.27),$$
$$a(r) = 2\pi\int F(H_r)J_n(2\pi r H_r)H_r\,\mathrm{d}H_r. \quad (4.2.5.28)$$

(2) *p-fold symmetry of the projected molecule*, where $a(r,\psi) = a[r,\psi+(2\pi/p)]$

$$F_p(H_r,\Psi) = \sum_n \exp\{inp[\Psi+(\pi/2)]\} \times \int a_{np}(r)J_{np}(2\pi r H_r)2\pi r\,\mathrm{d}r, \quad (4.2.5.29)$$

$$a_p(r,\psi) = \sum_n |a_{np}(r)|\cos[np\psi - \psi_{np}(r)]. \quad (4.2.5.30)$$

Only Bessel functions $J_0, J_p, J_{2p}, \ldots$ occur. In most cases J_{2p} and higher orders may be neglected.

(3) *Vertical mirror planes*

Only cosine terms occur, *i.e.* all $\beta_n = 0$ or $\psi_n(r) = 0$.

The *general 3D expressions* valid for extended chains with period c [equation (4.2.5.12)] are found in an analogous way,

$$a(r,\psi,z) = a_{\mathrm{M}}(r,\psi,z) * l(z),$$

$$F(\mathbf{H}) = F_l(H_r,\Psi,L) = F_{\mathrm{M}}(\mathbf{H})L(L) = \int\int\int a_{\mathrm{M}}(r,\psi,z)\exp\{2\pi i\,[H_r r\cos(\psi-\Psi)+Lz]\} \times 2\pi r\,\mathrm{d}r\,\mathrm{d}\psi\,\mathrm{d}z. \quad (4.2.5.31)$$

Using a series expansion analogous to (4.2.5.23) and (4.2.5.24),

$$a_{nl}(r) = \frac{1}{2\pi}\int\int a_{\mathrm{M}}\exp\{-i(n\psi - 2\pi lz)\}\,\mathrm{d}\psi\,\mathrm{d}z, \quad (4.2.5.32)$$

$$F_{nl}(H_r) = \exp\{in\pi/2\}\int a_{nl}(r)J_n(2\pi H_r r)2\pi r\,\mathrm{d}r, \quad (4.2.5.33)$$

one has

$$F_l(\mathbf{H}) = \sum_n \exp\{in[\Psi+(\pi/2)]\}\int a_{nl}(r)J_n(2\pi H_r r)2\pi r\,\mathrm{d}r. \quad (4.2.5.34)$$

In practice the integrals are often replaced by discrete summation of j atoms at positions $r = r_j$, $\psi = \psi_j$, $z = z_j$ ($0 \le z_j < c$):

$$F_l(\mathbf{H}) = \sum_j\sum_n f_j J_n(2\pi H_r r_j)\exp\{-in\psi_j\} \times \exp\{2\pi ilz_j\}\exp\{in[\Psi+(\pi/2)]\} \quad (4.2.5.35)$$

or

$$F_l(\mathbf{H}) = \sum_n(\alpha_n + i\beta_n)\exp\{in\Psi\},$$

$$\alpha_n = \sum_j f_j J_n(2\pi H_r r_j)\cos\{n[(\pi/2)-\psi_j] + 2\pi lz_j\},$$
$$\beta_n = \sum_j f_j J_n(2\pi H_r r_j)\sin\{n[(\pi/2)-\psi_j] + 2\pi lz_j\}.$$

Intensity in the lth diffuse layer is given by

$$I_l = \sum_n\sum_{n'}[(\alpha_n\alpha_{n'} + \beta_n\beta_{n'}) + i(\alpha_{n'}\beta_n - \alpha_n\beta_{n'})] \times \exp\{i(n-n')\Psi\}. \quad (4.2.5.36)$$

(*a*) *Cylinder symmetry* (free rotating molecules around the chain axis or statistical averaging with respect to ψ over an assembly of chains). Only component F_{0l} occurs:

$$F_{0l}(H_r, L) = 2\pi\int\int\langle a_{\mathrm{M}}\rangle J_0(2\pi H_r r)\exp\{2\pi ilz\}r\,\mathrm{d}r\,\mathrm{d}z$$

or

$$F_{0l}(H_r, L) = \sum_j f_j J_0(2\pi H_r r_j) \exp\{2\pi i l z_j\}.$$

In particular, $F_{00}(H_r)$ determines the radial component of the molecule projected along z:

$$F_{00}(H_r) = \sum_j f_j J_0(2\pi H_r r_j).$$

(*b*) *p-fold symmetry* of a plane molecule (or projected molecule) as outlined previously: only components *np* instead of *n* occur. Bessel functions J_0 and J_p are sufficient in most cases.

(*c*) *Vertical mirror plane*: see above.

(*d*) *Horizontal mirror plane* (perpendicular to the chain): Exponentials $\exp\{2\pi i l z\}$ in equation (4.2.5.32) may be replaced by $\cos 2\pi l z$.

(*e*) *Twofold symmetry axis* perpendicular to the chain axis (at positions $\psi = 0,\ 2\pi/p, \ldots$). Exponentials $\exp\{-i(np\psi - 2\pi l z)\}$ in equation (4.2.5.32) are replaced by the corresponding cosine term $\cos(np\psi + 2\pi l z)$.

Formulae concerning the reverse method (Fourier synthesis) are not given here (see, *e.g.*, Vainshtein, 1966). There is usually no practical use for this in diffuse-scattering work because it is very difficult to separate out a single component F_{nl}. *Every* diffuse layer is affected by *all* components F_{nl}. There is a chance of doing so only if *one* diffuse layer corresponds predominantly to *one* Bessel function.

4.2.5.3.2. Disorder within randomly distributed collinear chains

Deviations from strict periodicities in the z direction within one chain may be due to loss of translational symmetry of the centres of the molecules along z and/or due to varying orientations of the molecules with respect to different axes, such as azimuthal misorientation, tilting with respect to the z axis or combinations of both types. As in 3D crystals, there may or may not exist 1D structures in an averaged sense.

(1) *General treatment*

All formulae given in this section are only special cases of a 3D treatment (see, *e.g.*, Guinier, 1963). The 1D lattice (4.2.5.12) is replaced by a distribution:

$$\begin{aligned} d(z) &= \sum_\nu \delta(z - z_\nu), \\ D(L) &= \sum_\nu \exp\{2\pi i L z_\nu\}, \\ F(\mathbf{H}) &= F_{\rm M}(\mathbf{H}) D(L). \end{aligned} \tag{4.2.5.37}$$

The Patterson function is given by

$$P(\mathbf{r}) = [a_{\rm M}(\mathbf{r}) * a_{\rm M}(-\mathbf{r})] * [d(z) * d(-z)]. \tag{4.2.5.38}$$

Because the autocorrelation function $w = d * d$ is centrosymmetric,

$$w(z) = N\delta(z) + \sum_\nu \sum_\mu \delta[z - (z_\nu - z_\mu)] + \sum_\nu \sum_\mu \delta[z + (z_\nu - z_\mu)], \tag{4.2.5.39}$$

the interference function $W(L)$ $(= |D(L)|^2)$ is given by

$$W(L) = N + 2 \sum_\nu \sum_\mu \cos 2\pi [L(z_\nu - z_\mu)], \tag{4.2.5.40}$$

$$I(\mathbf{H}) = |F_{\rm M}(\mathbf{H})|^2 W(L). \tag{4.2.5.41}$$

Sometimes, *e.g.* in the following example of orientational disorder, there is an order only within domains. As shown in Section 4.2.3, this may be treated by a box or shape function $b(z) = 1$ for $z \le z_N$ and 0 elsewhere.

$$\begin{aligned} d(z) &= d_\infty b(z), \\ a(\mathbf{r}) &= a_{\rm M}(\mathbf{r}) * [d_\infty b(z)], \\ F(\mathbf{H}) &= F_{\rm M}(\mathbf{H})[D_\infty * B(L)], \end{aligned} \tag{4.2.5.42}$$

with

$$\begin{aligned} b(z) * b(-z) &\leftrightarrow |B(L)|^2, \\ I &= |F_{\rm M}(\mathbf{H})|^2 |D_\infty * B(L)|^2. \end{aligned} \tag{4.2.5.43}$$

If the order is perfect within one domain one has $D_\infty(L) \simeq \sum \delta(L - l)$; $(D_\infty * B) = \sum D(L - l)$; *i.e.* each reflection is affected by the shape function.

(2) *Orientational disorder*

A misorientation of the chain molecules with respect to one another is taken into account by different structure factors $F_{\rm M}$.

$$I(\mathbf{H}) = \sum_\nu \sum_\mu F_\nu(\mathbf{H}) F_\mu(\mathbf{H})^+ \exp\{2\pi i L(z_\nu - z_\mu)\}. \tag{4.2.5.44}$$

A further discussion follows the same arguments outlined in Section 4.2.3. For example, a very simple result is found in the case of uncorrelated orientations. Averaging over all pairs $F_\nu F_\mu^+$ yields

$$I(\mathbf{H}) = N(\langle |F|^2 \rangle - \langle |F| \rangle^2) + \langle |F| \rangle^2 L(L), \tag{4.2.5.44a}$$

where

$$\begin{aligned} \langle |F| \rangle^2 &= 1/N^2 \langle F_\nu F_\mu^+ \rangle \\ &= \sum_\nu \alpha_\nu F_\nu(\mathbf{H}) \sum_\mu \alpha_\mu F_\mu^+(\mathbf{H}), \quad (\nu \ne \mu) \\ \langle |F|^2 \rangle &= 1/N \langle F_\nu F_\nu^+ \rangle = \sum_\nu \alpha_\nu |F_\nu(\mathbf{H})|^2. \end{aligned}$$

Besides the diffuse layer system, there is a diffuse background modulated by the $\mathbf{H}$ dependence of $[\langle |F(\mathbf{H})|^2 \rangle - \langle |F(\mathbf{H})| \rangle^2]$.

(3) *Longitudinal disorder*

In this context, the structure factor of a chain molecule is neglected. Irregular distances between the molecules within a chain occur owing to the shape of the molecules, intrachain interactions and/or interaction forces *via* a surrounding matrix. A general discussion is given by Guinier (1963). It is convenient to reformulate the discrete Patterson function, *i.e.* the correlation function (4.2.5.39),

$$\begin{aligned} w(z) = N\delta(z) &+ \sum_\nu \delta[z \pm (z_\nu - z_{\nu+1})] \\ &+ \sum_\nu \delta[z \pm (z_\nu - z_{\nu+2})] + \ldots \end{aligned} \tag{4.2.5.39a}$$

in terms of continuous functions $a_\mu(z)$ that describe the probability of finding the μth neighbour within an arbitrary distance,

$$\begin{aligned} w'(z) = w(z)/N = \delta(z) &+ a_1(z) + a_{-1}(z) + \ldots \\ &+ a_\mu(z) + a_{-\mu}(z) + \ldots \end{aligned} \tag{4.2.5.45}$$

$[\int a_\mu(z)\,{\rm d}z = 1,\ a_\mu(z) = a_{-\mu}(-z)]$.

There are two principal ways to define $a_\mu(z)$. The first is the case of a well defined one-dimensional lattice with positional fluctuations of the molecules around the lattice points, *i.e.* long-range order is retained: $a_\mu(z) = \mu c_0 + z_\mu$, where z_μ denotes the displacement of the μth molecule in the chain. Gaussian distributions are frequently used:

$$c' \exp\{-(z - \mu c_0)^2 / 2\Delta^2\},$$

where c' is a normalizing constant and Δ is the standard deviation. Fourier transformation of equation (4.2.5.45) gives the well known result

$$I_D \simeq (1 - \exp\{-L^2\Delta^2\}),$$

i.e. a monotonically increasing intensity with L (any modulation due to a molecular structure factor is neglected). This result is quite analogous to the treatment of the scattering of independently vibrating atoms. If (short-range) correlations exist between the molecules, the Gaussian distribution is replaced by a multivariate normal distribution where correlation coefficients κ^μ $(0 < \kappa < 1)$ between a molecule and its μth neighbour are incorporated. κ^μ is defined by the second moment: $\langle z_0 z_\mu \rangle / \Delta^2$.

$$a_\mu(z) = c'' \exp\{-(z - \mu c_0)^2 / 2\Delta^2(1 - \kappa^\mu)\}.$$

Obviously, the variance increases if the correlation diminishes and reaches an upper bound of twice the single-site variance. Fourier transformation gives an expression for the diffuse intensity (Welberry, 1985):

$$I_D(L) \simeq \exp\{-L^2\Delta^2\} \sum_j (-L^2\Delta^2)^j / j! \times (1 - \kappa^{2j})/(1 + \kappa^{2j} - 2\kappa^j \cos 2\pi L c_0). \quad (4.2.5.46)$$

For small Δ, terms with $j > 1$ are mostly neglected. The terms become increasingly important with higher values of L. On the other hand, κ^j becomes smaller with increasing j, each additional term in equation (4.2.5.46) becomes broader and, as a consequence, the diffuse planes in reciprocal space become broader with higher L.

In a different way – in the paracrystal method – the position of the second and subsequent molecules with respect to some reference zero point depends on the *actual* position of the predecessor. The variance of the position of the μth molecule relative to the first becomes unlimited. There is a continuous transition to a fluid-like behaviour of the chain molecules. This 1D paracrystal (these are sometimes called distortions of the second kind) is only a special case of the 3D paracrystal concept (see Hosemann & Bagchi, 1962; Wilke, 1983). Despite some difficulties with this concept (Brämer, 1975; Brämer & Ruland, 1976), it is widely used as a theoretical model for describing diffraction by highly distorted lattices. One essential development is to limit the size of a paracrystalline grain so that fluctuations never become too large (Hosemann, 1975).

If this concept is used for the 1D case, $a_\mu(z)$ is defined by convolution products of $a_1(z)$. For example, the probability of finding the next-nearest molecule is given by

$$a_2(z) = \int a_1(z')\, a_1(z - z')\, \mathrm{d}z' = a_1(z) * a_1(z)$$

and, generally,

$$a_\mu(z) = a_1(z) * a_1(z) * a_1(z) * \ldots * a_1(z)$$

(μ-fold convolution).

The mean distance between next-nearest neighbours is

$$\langle c \rangle = \int z' a_1(z')\, \mathrm{d}z'$$

and between neighbours of the μth order it is $\mu\langle c \rangle$. The average value of $a_\mu = 1/\langle c \rangle$, which is also the value of $w(z)$ for $z > z_k$, where the distribution function is completely smeared out. The general expression for the interference function $G(L)$ is

$$G(L) = 1 + \sum_\mu \{F^\mu + F^{+\mu}\} = \mathrm{Re}\{(1 + F)/(1 - F)\} \quad (4.2.5.47)$$

with $F(L) \leftrightarrow a_1(z)$, $F^\mu(L) \leftrightarrow a_\mu(z)$.

With $F = |F| \exp\{i\chi\}$ $(\chi = L\langle c \rangle)$, equation (4.2.5.47) is written

$$G(L) = [1 - |F(L)|^2]/[1 - 2|F(L)| \cos\chi + |F(L)|^2]. \quad (4.2.5.47a)$$

[Note the close similarity to the diffuse part of equation (4.2.5.5), which is valid for 1D disorder problems.]

This function has maxima of height $(1 + |F|)/(1 - |F|)$ and minima of height $(1 - |F|)/(1 + |F|)$ at positions lc^* and $(l + \frac{1}{2})c^*$, respectively. With decreasing $|F|$ the oscillations vanish; a critical L value (corresponding to z_k) may be defined by $G_{max}/G_{min} \lesssim 1.2$. Actual values depend strongly on $F(\mathbf{H})$.

The paracrystal method is substantiated by the choice $a_1(z)$, *i.e.* the disorder model. Again, a Gaussian distribution is frequently used:

$$a_1(z) = (1/\sqrt{2\pi}\Delta) \exp\{-(z - \langle c \rangle)^2 / 2\Delta^2\}$$
$$a_\mu(z) = (1/\sqrt{\mu})(1/\sqrt{2\pi}\Delta) \exp\{-(z - \mu\langle c \rangle)^2 / 2\mu\Delta^2\} \quad (4.2.5.48)$$

with the two parameters $\langle c \rangle$, Δ.

There are peaks of height $1/[\pi^2 L^2 (\Delta/\langle c \rangle)^2]$ which obviously decrease with L^2 and $(\Delta/\langle c \rangle)^2$. The oscillations vanish for $|F| \simeq 0.1$, *i.e.* $1/\langle c \rangle \simeq 0.25/\Delta$. The width of the mth peak is $\Delta_m = \sqrt{m}\Delta$. The integral reflectivity is approximately $1/\langle c \rangle [1 - \pi^2 L^2 (\Delta/\langle c \rangle)^2]$ and the integral width (defined by integral reflectivity divided by peak reflectivity) (background subtracted!) is $1/\langle c \rangle \pi^2 L^2 (\Delta/\langle c \rangle)^2$ which, therefore, increases with L^2. In principle, the same results are given by Zernike & Prins (1927). In practice, a single Gaussian distribution is not fully adequate and modified functions must be used (Rosshirt *et al.*, 1985).

Our final remark concerns the normalization [equation (4.2.5.39)]. Going from (4.2.5.39) to (4.2.5.45) it is assumed that N is a large number so that the correct normalization factors $(N - |\mu|)$ for each $a_\mu(z)$ may be approximated by a uniform N. If this is not true then

$$G(L) = N + \sum_\mu (N - |\mu|)(F^\mu + F^{+\mu}) = N\,\mathrm{Re}\{(1 + |F|)/(1 - |F|)\} - 2\,\mathrm{Re}\{|F|(1 - |F|^N)/(1 - |F|^2)\}. \quad (4.2.5.49)$$

The correction term may be important in the case of relatively small (1D) domains. As mentioned above, the structure factor of a chain molecule was neglected. The $\mathbf{H}$ dependence of F_M, of course, obscures the intensity variation of the diffuse layers as described by (4.2.5.47*a*).

The matrix method developed for the case of planar disorder was adapted to 2D disorder by Scaringe & Ibers (1979). Other models and corresponding expressions for diffuse scattering are developed from specific microscopic models (potentials), *e.g.* in the case of $Hg_{3-\delta}AsF_6$ (Emery & Axe, 1978; Radons *et al.*, 1983), hollandites (Beyeler *et al.*, 1980; Ishii, 1983), iodine chain compounds (Endres *et al.*, 1982) or urea inclusion compounds (Forst *et al.*, 1987; Weber *et al.*, 2000).

4.2.5.3.3. Correlations between almost collinear chains

In real cases there are quite strong correlations between different chains, at least within small domains. Deviations from a strict (3D) order of chain-like structural elements are due to several reasons: the shape and structure of the chains, varying binding forces, and thermodynamical or kinetic considerations.

Many types of disorder occur. (1) Relative shifts parallel to the common axis while projections along this axis give a perfect 2D ordered net ('axial disorder'). (2) Relative fluctuations of the distances between the chains (perpendicular to the unique axis) with short-range order along the transverse **a** and/or **b** directions. The net of projected chains down to the *ab* plane is distorted ('net distortions'). Disorder of types (1) and (2) is sometimes correlated owing to nonuniform cross sections of the chains. (3) Turns, twists and torsions of chains or parts of chains. This azimuthal type of disorder may be treated in a similar way to the case of

azimuthal disorder of single-chain molecules. Correlations between axial shifts and torsions produce 'screw shifts' (helical structures). Torsion of chain parts may be of dynamic origin (such as rotational vibrations). (4) Tilting or bending of the chains in a uniform or nonuniform way ('conforming/nonconforming' disorder). Many of these types and a variety of combinations between them are found in polymer and liquid crystals, and are therefore treated separately. Only some simple basic ideas are discussed here in brief.

For the sake of simplicity, the paracrystal concept in combination with Gaussians is used again. Distribution functions are given by convolution products of next-nearest-neighbour distribution functions. As long as averaged lattice directions and lattice constants in a plane perpendicular to the chain axis exist, only two functions, $a_{100} = a_1(xyz)$ and $a_{010} = a_2(xyz)$, are needed to describe the arrangement of next-nearest chains. Longitudinal disorder is treated as before by a third distribution function $a_{001} = a_3(xyz)$. The phenomena of chain bending or tilting may be incorporated by an x and y dependence of a_3. Any general fluctuation in the spatial arrangement of chains is given by

$$a_{mpq} = a_1 * \ldots * a_1 * a_2 * \ldots * a_2 * a_3 * \ldots * a_3. \qquad (4.2.5.50)$$

(m-fold, p-fold, q-fold self-convolution of a_1, a_2, a_3, respectively.)

$$w(\mathbf{r}) = \delta(\mathbf{r}) \sum_m \sum_p \sum_q [a_{mpq}(\mathbf{r}) + a_{-mpq}(\mathbf{r})]. \qquad (4.2.5.51)$$

a_ν ($\nu = 1, 2, 3$) are called fundamental functions. If an averaged lattice cannot be defined, more fundamental functions a_ν are needed to account for correlations between them.

By Fourier transformations, the interference function is given by

$$G(\mathbf{H})\, G_\nu = \mathrm{Re}\{(1 + |F_\nu|)/(1 - |F_\nu|)\}. \qquad (4.2.5.52)$$

If Gaussian functions are assumed, simple pictures are derived. For example

$$a_1(\mathbf{r} + \langle \mathbf{a} \rangle) = [1/(2\pi)^{3/2}][1/(\Delta_{11}\Delta_{12}\Delta_{13})] \times \exp\left\{-\tfrac{1}{2}[(x^2/\Delta_{11}^2) + (y^2/\Delta_{12}^2) + (z^2/\Delta_{13}^2)]\right\} \qquad (4.2.5.53)$$

describes the distribution of neighbours in the x direction (with a mean distance $\langle a \rangle$). Parameters Δ_{13}, Δ_{11} and Δ_{12} concern axial, radial and tangential fluctuations, respectively. Pure axial distribution along **c** is given by projection of a_1 on the z axis; pure net distortions are given by projection on the xy plane. If the chain-like structure is neglected, the interference function

$$G_1(\mathbf{H}) = \exp\{-2\pi^2(\Delta_{11}^2 H^2 + \Delta_{12}^2 K^2 + \Delta_{13}^2 L^2)\} \qquad (4.2.5.54)$$

describes a set of diffuse planes perpendicular to $\mathbf{a}^*$ with a mean distance of $1/\langle a \rangle$. These diffuse layers broaden along H with $m\Delta_{11}$ and decrease in intensity along K and L monotonically. There is an ellipsoid-shaped region in reciprocal space defined by main axes of length $1/\Delta_{11}, 1/\Delta_{12}, 1/\Delta_{13}$ with a limiting surface given by $|F| \simeq 0.1$, beyond which the diffuse intensity is completely smeared out. The influence of a_2 may be discussed in an analogous way.

If the chain-like arrangement parallel to **c** [equation (4.2.5.12)] is taken into consideration,

$$l(z) = \sum_{n_3} \delta(z - n_3 c),$$

the set of planes perpendicular to $\mathbf{a}^*$ (and/or $\mathbf{b}^*$) is subdivided in the L direction by a set of planes located at $l(1/c)$ [equation (4.2.5.15)].

Longitudinal disorder is given by $a_3(z)$ [equation (4.2.5.48), $\Delta_{33} = \Delta$] and leads to two intersecting sets of broadened diffuse layer systems.

Particular cases like pure axial distributions ($\Delta_{11}, \Delta_{12} \simeq 0$), pure tangential distributions (net distortions: $\Delta_{11}, \Delta_{13} \simeq 0$), uniform bending of chains or combinations of these effects are discussed in the monograph by Vainshtein (1966).

4.2.5.4. Defects, short-range ordering, clustering: three-dimensional disorder

4.2.5.4.1. General formulation (elastic diffuse scattering)

In this section general formulae for diffuse scattering will be derived that may best be applied to crystals with a well ordered average structure, characterized by (almost) sharp Bragg peaks. Textbooks and review articles concerning defects and local ordering are by Krivoglaz (1969, 1996*a*,*b*), Dederichs (1973), Peisl (1975), Schwartz & Cohen (1977), Schmatz (1973, 1983), Bauer (1979), Kitaigorodsky (1984) and Schweika (1998). A series of interesting papers on local order is given by Young (1975) and also by Cowley *et al.* (1979). Expressions for polycrystalline samples are given by Warren (1969) and Fender (1973).

Two general methods may be applied: (*a*) the average difference cluster method, where a representative cluster of scattering differences between the average structure and the cluster is used; and (*b*) the method of short-range-order correlation functions, where formal parameters are introduced.

The two methods are equivalent in principle. The cluster method is generally more convenient in cases where a single average cluster is a good approximation. This holds for small concentrations of clusters. In the literature this problem is treated in terms of fluctuations of the distribution functions as will be discussed below (Section 4.2.5.4.6). The most convenient way to derive the distribution function correctly from experimental data is the use of low-angle scattering, which generally shows one or more clear maxima caused by partly periodic properties of the distribution function. For the deconvolution of the distribution function, achieved by Fourier transformation of the corrected diffuse low-angle scattering, the reader is referred to the relevant literature. However, deconvolutions are not unique and some reasonable assumptions are necessary for a final solution.

The method of short-range-order parameters is optimal in cases where isolated clusters are not realized and the correlations do not extend to long distances. Otherwise periodic solutions are more convenient in most cases.

In any case, the first step towards the solution of the diffraction problem is the accurate determination of the average structure. As described in Section 4.2.4, important information on fractional occupations, interstitials and displacements of atoms (shown by unusual thermal parameters) may be derived. Unfortunately, all defects contribute to diffuse scattering; hence one has to start with the assumption that the disorder to be interpreted is predominant. Fractional occupancy of certain lattice sites by two or more kinds of atoms plays an important role in the literature, especially in metallic or ionic structures. Since vacancies may be treated as atoms with zero scattering amplitude, structures containing vacancies may be formally treated as multicomponent systems.

Since the solution of the diffraction problem should not be restricted to metallic systems with a simple (primitive) structure, we have to consider the structure of the unit cell – as given by the average structure – and the propagation of order according to the translation group separately. In simple metallic systems this difference is immaterial. It is well known that the thermodynamic problem of propagation of order in a three-dimensional crystal can hardly ever be solved analytically in a general way. Some solutions have been published with the aid of the Ising model using next-nearest-neighbour interactions. They are excellent for

an understanding of the principles of order–disorder phenomena, but they can rarely be applied quantitatively in practical problems. Hence, methods have been developed to derive the propagation of order from the diffraction pattern by means of Fourier transformation. This method has been described qualitatively in Section 4.2.3.1 and will be used here for a quantitative application. In a first approximation, the assumption of a small number of different configurations of the unit cell is made, represented by the corresponding number of structure factors. Displacements of atoms caused by the configurations of the neighbouring cells are excluded. This problem will be treated subsequently.

The finite number of structures of the unit cell in the disordered crystal is given by

$$F_\nu(\mathbf{r}) = \sum_j \sum_\mu \pi^\nu_{j\mu} f_\mu(\mathbf{r} - \mathbf{r}_j). \tag{4.2.5.55}$$

Note that $F_\nu(\mathbf{r})$ is defined in real space and $\mathbf{r}_j$ gives the position vector of site j; $\pi^\nu_{j\mu} = 1$ if in the νth structure factor the site j is occupied by an atom of kind μ, and 0 elsewhere.

In order to apply the laws of Fourier transformation adequately, it is useful to introduce the distribution function of F_ν,

$$\pi_\nu(\mathbf{r}) = \sum_{\mathbf{n}} \pi_{\mathbf{n}\nu} \delta(\mathbf{r} - \mathbf{n}), \tag{4.2.5.56}$$

with $\pi_{\mathbf{n}\nu} = 1$ if the cell $\mathbf{n} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$ has the F_ν structure and $\pi_{\mathbf{n}\nu} = 0$ elsewhere.

In the definitions given above $\pi_{\mathbf{n}\nu}$ are numbers (scalars) assigned to the cell. Since all these are occupied we have

$$\sum_\nu \pi_\nu(\mathbf{r}) = l(\mathbf{r})$$

where $l(\mathbf{r})$ is the lattice in real space.

The structure of the disordered crystal is given by

$$\sum_\nu \pi_\nu(\mathbf{r}) * F_\nu(\mathbf{r}). \tag{4.2.5.57}$$

$\pi_\nu(\mathbf{r})$ consists of $\alpha_\nu N$ points, where $N = N_1N_2N_3$ is the total (large) number of unit cells and α_ν denotes the *a priori* probability (concentration) of the νth cell occupation.

It is now useful to introduce

$$\Delta\pi_\nu(\mathbf{r}) = \pi_\nu(\mathbf{r}) - \alpha_\nu l(\mathbf{r}) \tag{4.2.5.58}$$

with

$$\sum_\nu \Delta\pi_\nu(\mathbf{r}) = \sum_\nu \pi_\nu(\mathbf{r}) - l(\mathbf{r}) \sum_\nu \alpha_\nu = l(\mathbf{r}) - l(\mathbf{r}) = 0.$$

Introducing (4.2.5.58) into (4.2.5.57) gives

$$\begin{aligned}\sum_\nu \pi_\nu(\mathbf{r}) * F_\nu(\mathbf{r}) &= \sum_\nu \Delta\pi_\nu(\mathbf{r}) * F_\nu(\mathbf{r}) + l(\mathbf{r}) \sum_\nu \alpha_\nu F_\nu(\mathbf{r}) \\ &= \sum_\nu \Delta\pi_\nu(\mathbf{r}) * F_\nu(\mathbf{r}) + l(\mathbf{r})\langle F(\mathbf{r})\rangle.\end{aligned} \tag{4.2.5.59}$$

Similarly

$$\begin{aligned}\Delta F_\nu(\mathbf{r}) &= F_\nu(\mathbf{r}) - \langle F(\mathbf{r})\rangle, \\ \sum_\nu \alpha_\nu \Delta F_\nu(\mathbf{r}) &= \sum_\nu \alpha_\nu F_\nu(\mathbf{r}) - \sum_\nu \alpha_\nu \langle F(\mathbf{r})\rangle = 0.\end{aligned} \tag{4.2.5.60}$$

Using (4.2.5.60) it follows from (4.2.5.58) that

$$\begin{aligned}\sum_\nu \pi_\nu(\mathbf{r}) * F_\nu(\mathbf{r}) &= \sum_\nu \Delta\pi_\nu(\mathbf{r}) * \Delta F_\nu(\mathbf{r}) \\ &\quad + \sum_\nu \Delta\pi_\nu(\mathbf{r}) * \langle F(\mathbf{r})\rangle + l(\mathbf{r}) * \langle F(\mathbf{r})\rangle \\ &= \sum_\nu \Delta\pi_\nu(\mathbf{r}) * \Delta F_\nu(\mathbf{r}) + l(\mathbf{r}) * \langle F(\mathbf{r})\rangle.\end{aligned} \tag{4.2.5.61}$$

Comparison with (4.2.5.59) yields

$$\sum_\nu \Delta\pi_\nu(\mathbf{r}) * F_\nu(\mathbf{r}) = \sum_\nu \Delta\pi_\nu(\mathbf{r}) * \Delta F_\nu(\mathbf{r}).$$

Fourier transformation of (4.2.5.61) gives

$$\sum_\nu \Pi_\nu(\mathbf{H}) F_\nu(\mathbf{H}) = \sum_\nu \Delta\Pi_\nu(\mathbf{H}) \Delta F_\nu(\mathbf{H}) + L(\mathbf{H})\langle F(\mathbf{H})\rangle$$

with

$$\sum_\nu \Delta\Pi_\nu(\mathbf{H}) = 0, \quad \sum_\nu \Delta F_\nu(\mathbf{H}) = 0.$$

The expression for the scattered intensity is therefore

$$\begin{aligned}I(\mathbf{H}) &= \left|\sum_\nu \Delta\Pi_\nu(\mathbf{H}) \Delta F_\nu(\mathbf{H})\right|^2 + |L(\mathbf{H})\langle F(\mathbf{H})\rangle|^2 \\ &\quad + L(\mathbf{H})\Bigg\{\langle F^+(\mathbf{H})\rangle \sum_\nu \Delta\Pi_\nu(\mathbf{H}) F_\nu(\mathbf{H}) \\ &\quad + \langle F^+(\mathbf{H})\rangle \sum_\nu \Delta\Pi^+_\nu(\mathbf{H}) F^+_\nu(\mathbf{H})\Bigg\}.\end{aligned} \tag{4.2.5.62}$$

Because of the multiplication by $L(\mathbf{H})$, the third term in (4.2.5.62) contributes to sharp reflections only. Since they are correctly given by the second term in (4.2.5.62), the third term vanishes. Hence the diffuse part is given by

$$I_{\rm D}(\mathbf{H}) = \left|\sum_\nu \Delta\Pi_\nu(\mathbf{H}) \Delta F_\nu(\mathbf{H})\right|^2. \tag{4.2.5.63}$$

For a better understanding of the behaviour of diffuse scattering it is useful to return to real space:

$$\begin{aligned}i_{\rm D}(\mathbf{r}) &= \sum_\nu \Delta\pi_\nu(\mathbf{r}) * \Delta F_\nu(\mathbf{r}) * \sum_{\nu'} \Delta\pi_{\nu'}(-\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r}) \\ &= \sum_\nu \sum_{\nu'} \Delta\pi_\nu(\mathbf{r}) * \Delta\pi_{\nu'}(-\mathbf{r}) * \Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r})\end{aligned} \tag{4.2.5.64}$$

and with (4.2.5.58):

$$\begin{aligned}i_{\rm D}(\mathbf{r}) &= \sum_\nu \sum_{\nu'} [\pi_\nu(\mathbf{r}) - \alpha_\nu l(\mathbf{r})] * [\pi_{\nu'}(-\mathbf{r}) - \alpha_{\nu'} l(\mathbf{r})] \\ &\quad * \Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r}).\end{aligned} \tag{4.2.5.65}$$

Evaluation of this equation for a single term yields

$$\begin{aligned}&[\pi_\nu(\mathbf{r}) * \pi_{\nu'}(-\mathbf{r}) - \alpha_\nu l(\mathbf{r}) * \pi_{\nu'}(-\mathbf{r}) \\ &\quad - \alpha_{\nu'} l(\mathbf{r}) * \pi_\nu(\mathbf{r}) + \alpha_\nu \alpha_{\nu'} l(\mathbf{r}) * l(\mathbf{r})] \\ &\quad * \Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(\mathbf{r}).\end{aligned} \tag{4.2.5.66}$$

Since $l(\mathbf{r})$ is a periodic function of points, all convolution products with $l(\mathbf{r})$ are also periodic. For the final evaluation the decrease of the number of overlapping points (maximum N) in the convolution products with increasing displacements of the functions is neglected (it is assumed that there is no particle-size effect). Then (4.2.5.66) becomes

$$[\pi_\nu(\mathbf{r}) * \pi_{\nu'}(-\mathbf{r}) - N\alpha_\nu\alpha_{\nu'}l(\mathbf{r}) - N\alpha_\nu\alpha_{\nu'}l(\mathbf{r}) + N\alpha_\nu\alpha_{\nu'}l(\mathbf{r})]$$
$$* \Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r})$$
$$= [\pi_\nu(\mathbf{r}) * \pi_{\nu'}(-\mathbf{r}) - N\alpha_\nu\alpha_{\nu'}l(\mathbf{r})]$$
$$* \Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r}). \quad (4.2.5.67)$$

If the first term in (4.2.5.67) is considered, the convolution of the two functions for a given distance $\mathbf{n}$ counts the number of coincidences of the function $\pi_\nu(\mathbf{r})$ with $\pi_{\nu'}(-\mathbf{r})$. This quantity is given by $Nl(\mathbf{r})\alpha_\nu p_{\nu\nu'}(-\mathbf{r})$, where $\alpha_\nu p_{\nu\nu'}(\mathbf{r})$ is the probability of a pair occupation in the $\mathbf{r}$ direction.

Equation (4.2.5.67) then reads

$$[Nl(\mathbf{r})\alpha_\nu p_{\nu\nu'}(-\mathbf{r}) - N\alpha_\nu\alpha_{\nu'}l(\mathbf{r})]$$
$$* \Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r})$$
$$= N\alpha_\nu[p'_{\nu\nu'}(-\mathbf{r})l(\mathbf{r})]$$
$$* [\Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r})] \quad (4.2.5.68)$$

with $p'_{\nu\nu'}(\mathbf{r}) = p_{\nu\nu'}(\mathbf{r}) - \alpha_{\nu'}$. The function $\alpha_\nu p'_{\nu\nu'}(\mathbf{r})$ is usually called the pair-correlation function $g \leftrightarrow \alpha_{\nu\nu'|\mathbf{nn}'}$ in physics.

The following relations hold:

$$\sum_\nu \alpha_\nu = 1; \quad (4.2.5.69a)$$
$$\sum_{\nu'} \alpha_\nu p_{\nu\nu'}(\mathbf{r}) = \alpha_{\nu'}; \quad \sum_{\nu'} \alpha_\nu p'_{\nu\nu'}(\mathbf{r}) = 0 \quad (4.2.5.69b)$$
$$\sum_\nu \alpha_\nu p_{\nu\nu'}(\mathbf{r}) = \alpha_{\nu'}; \quad \sum_\nu \alpha_\nu p'_{\nu\nu'}(\mathbf{r}) = 0 \quad (4.2.5.69c)$$
$$\alpha_\nu p_{\nu\nu'}(\mathbf{r}) = \alpha_{\nu'} p_{\nu'\nu}(-\mathbf{r}). \quad (4.2.5.69d)$$

Functions normalized to unity are also in use. Obviously the following relation is valid: $p'_{\nu\nu'}(0) = \delta_{\nu\nu'} - \alpha_{\nu'}$.

Hence

$$\alpha_{\nu\nu'|\mathbf{nn}'} = \alpha_{\nu'} p'_{\nu\nu'}(\mathbf{r})/(\delta_{\nu\nu'} - \alpha_{\nu'})$$

is unity for $\mathbf{r} = 0\,(\mathbf{n} = \mathbf{n}')$. This property is especially convenient in binary systems.

With (4.2.5.68), equation (4.2.5.64) becomes

$$i_{\mathrm{D}}(\mathbf{r}) = N \sum_\nu \sum_{\nu'} [\alpha_\nu p'_{\nu\nu'}(-\mathbf{r})l(\mathbf{r})] * [\Delta F_\nu(\mathbf{r}) * \Delta F_{\nu'}(-\mathbf{r})] \quad (4.2.5.70)$$

and Fourier transformation yields

$$I_{\mathrm{D}}(\mathbf{H}) = N \sum_\nu \sum_{\nu'} [\alpha_\nu P'_{\nu\nu'}(-\mathbf{H}) * L(\mathbf{H})] \Delta F_\nu(\mathbf{H}) \Delta F^+_{\nu'}(\mathbf{H}). \quad (4.2.5.71)$$

It may be concluded from equations (4.2.5.69) that all functions $p'_{\nu\nu'}(\mathbf{r})$ may be expressed by $p'_{11}(\mathbf{r})$ in the case of two structure factors F_1, F_2. Then all $p'_{\nu\nu'}(\mathbf{r})$ are symmetric in $\mathbf{r}$; the same is true for the $P'_{\nu\nu'}(\mathbf{H})$. Consequently, the diffuse reflections described by (4.2.5.71) are all symmetric. The position of the diffuse peak depends strongly on the behaviour of $p'_{\nu\nu'}(\mathbf{r})$; in the case of cluster formation Bragg peaks and diffuse peaks coincide. Diffuse superstructure reflections are observed if the $p'_{\nu\nu'}(\mathbf{r})$ show some damped periodicities.

It should be emphasized that the condition $p'_{\nu\nu'}(\mathbf{r}) = p'_{\nu\nu'}(-\mathbf{r})$ may be violated for $\nu \neq \nu'$ if more than two cell occupations are involved. As shown below, the possibly asymmetric functions may be split into symmetric and antisymmetric parts. From equation (4.2.3.8) it follows that the Fourier transform of the antisymmetric part of $p'_{\nu\nu'}(\mathbf{r})$ is also antisymmetric. Hence, the convolution in the two terms in square brackets in (4.2.5.71) yields an antisymmetric contribution to each diffuse peak, generated by the convolution with the reciprocal lattice $L(\mathbf{h})$.

Obviously, equation (4.2.5.71) may also be applied to primitive lattices occupied by two or more kinds of atoms. Then the structure factors F_ν are merely replaced by the atomic scattering factors f_ν and the α_ν are equivalent to the concentrations of atoms c_ν. In terms of the $\alpha_{\nu\nu'|\mathbf{nn}'}$ (Warren short-range-order parameters) equation (4.2.5.71) reads

$$I_{\mathrm{D}}(\mathbf{H}) = N(\overline{f^2} - \bar{f}^2) \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \alpha_{\nu\nu'|\mathbf{nn}'} \exp\{2\pi i \mathbf{H} \cdot (\mathbf{n} - \mathbf{n}')\}. \quad (4.2.5.71a)$$

In the simplest case of a binary system A, B,

$$\alpha_{\mathbf{nn}'} = (1 - p_{AB|\mathbf{nn}'})/c_B = (1 - p_{BA|\mathbf{nn}'})/c_A;$$
$$c_A p_{AB} = c_B p_{BA}; \quad p_{AA} = 1 - p_{AB};$$

$$I_{\mathrm{D}}(\mathbf{H}) = N c_A c_B (f_A - f_B)^2 \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \alpha_{\mathbf{nn}'} \exp\{2\pi i \mathbf{H} \cdot (\mathbf{n} - \mathbf{n}')\}. \quad (4.2.5.71b)$$

[The exponential in (4.2.5.71*b*) may even be replaced by a cosine term owing to the centrosymmetry of this particular case.]

It should be mentioned that the formulations of the problem in terms of pair probabilities, pair-correlation functions, short-range-order parameters or concentration waves (Krivoglaz, 1969, 1996*a*,*b*) are equivalent. The Patterson function may also be used when using continuous electron (or nuclear) density functions where site occupancies are implied (Cowley, 1981).

4.2.5.4.2. Random distribution of defects

As shown above in the case of random distributions, all $p'_{\nu\nu'}(\mathbf{r})$ are zero except for $\mathbf{r} = 0$. Consequently, $p'_{\nu\nu'}(\mathbf{r})l(\mathbf{r})$ may be replaced by

$$\alpha_\nu p'_{\nu\nu'}(\mathbf{r}) = \alpha_\nu \delta_{\nu\nu'} - \alpha_\nu\alpha_{\nu'}. \quad (4.2.5.72)$$

According to (4.2.5.59) and (4.2.5.61), the diffuse scattering can be given by the Fourier transformation of

$$\sum_\nu \sum_{\nu'} \Delta\pi_\nu(\mathbf{r}) * \Delta\pi_{\nu'}(-\mathbf{r}) * F_\nu(\mathbf{r}) * F_\nu(-\mathbf{r})$$
$$= \sum_\nu \sum_{\nu'} p'_{\nu\nu'}(\mathbf{r}) * F_\nu(\mathbf{r}) * F_{\nu'}(-\mathbf{r})$$

or with (4.2.5.72)

$$i_{\mathrm{D}}(\mathbf{r}) = N \sum_\nu \sum_{\nu'} [\alpha_\nu \delta_{\nu\nu'} - \alpha_\nu\alpha_{\nu'}] * F_\nu(\mathbf{r}) * F_{\nu'}(-\mathbf{r}).$$

Fourier transformation gives

$$I_{\mathrm{D}}(\mathbf{H}) = N\left\{\sum_\nu \alpha_\nu |F_\nu(\mathbf{H})|^2 - \sum_\nu \alpha_\nu F_\nu(\mathbf{H}) \sum_{\nu'} \alpha_{\nu'} F^+_{\nu'}(\mathbf{H})\right\}$$
$$= N\{\langle |F(\mathbf{H})|^2 \rangle - |\langle F(\mathbf{H}) \rangle|^2\}. \quad (4.2.5.73)$$

This is the most general form of any diffuse scattering by systems ordered randomly ('Laue scattering'). Occasionally it is called 'incoherent scattering' (see Section 4.2.2).

4.2.5.4.3. Short-range order in multicomponent systems

The diffuse scattering of a disordered binary system without displacements of the atoms has already been discussed in Section 4.2.5.4.1. It could be shown that all distribution functions $p'_{\nu\nu'}(\mathbf{r})$ are mutually dependent and may be replaced by a single function [*cf.* (4.2.5.69)]. In that case $p'_{\nu\nu'}(\mathbf{r}) = p'_{\nu\nu'}(-\mathbf{r})$ was valid for all. This condition, however, may be violated in multicomponent systems. If a tendency towards an $F_1F_2F_3$ order in a ternary system is assumed, for example, $p_{12}(\mathbf{r})$ is apparently different from $p_{12}(-\mathbf{r})$. In this particular case it is useful to introduce

$$\langle p'_{\nu\nu'}(\mathbf{r})\rangle = \tfrac{1}{2}[p'_{\nu\nu'}(\mathbf{r}) + p'_{\nu\nu'}(-\mathbf{r})],$$
$$\Delta p'_{\nu\nu'}(\mathbf{r}) = \tfrac{1}{2}[p'_{\nu\nu'}(\mathbf{r}) - p'_{\nu\nu'}(-\mathbf{r})]$$

and their Fourier transforms $\langle P'_{\nu\nu'}(\mathbf{H})\rangle$, $\Delta P'_{\nu\nu'}(\mathbf{H})$, respectively.

The asymmetric correlation functions are therefore expressed by

$$p'_{\nu\nu'}(\mathbf{r}) = \langle p'_{\nu\nu'}(\mathbf{r})\rangle + \Delta p'_{\nu\nu'}(\mathbf{r});$$
$$p'_{\nu\nu'}(-\mathbf{r}) = \langle p'_{\nu\nu'}(\mathbf{r})\rangle - \Delta p'_{\nu\nu'}(\mathbf{r});$$
$$\Delta p'_{\nu\nu}(\mathbf{r}) = 0.$$

Consequently, $i_{\mathrm{D}}(\mathbf{r})$ (4.2.5.70) and $I_{\mathrm{D}}(\mathbf{H})$ (4.2.5.71) may be separated according to the symmetric and antisymmetric contributions. The final result is

$$\begin{aligned} I_{\mathrm{D}}(\mathbf{H}) = N\Big\{ & \sum_{\nu} \alpha_\nu [P'_{\nu\nu}(\mathbf{H}) * L(\mathbf{H})] |F_\nu(\mathbf{H})|^2 \\ & + \sum_{\nu>\nu'} \alpha_\nu [\langle P'_{\nu\nu'}(\mathbf{H})\rangle * L(\mathbf{H})] \\ & \times [\Delta F_\nu(\mathbf{H}) \Delta F^+_{\nu'}(\mathbf{H}) + \Delta F^+_\nu(\mathbf{H}) \Delta F_{\nu'}(\mathbf{H})] \\ & + \sum_{\nu>\nu'} \alpha_\nu [\Delta P'_{\nu\nu'}(-\mathbf{H}) * L(\mathbf{H})] \\ & \times [\Delta F_\nu(\mathbf{H}) \Delta F^+_{\nu'}(\mathbf{H}) - \Delta F^+_\nu(\mathbf{H}) \Delta F_{\nu'}(\mathbf{H})] \Big\}. \end{aligned} \tag{4.2.5.74}$$

Obviously, antisymmetric contributions to line profiles will only occur if structure factors of acentric cell occupations are involved. This important property may be used to draw conclusions with respect to the structure factors involved in the statistics. It should be mentioned here that the Fourier transform of the antisymmetric function $\Delta p'_{\nu\nu'}(\mathbf{r})$ is imaginary and antisymmetric. Since the last term in (4.2.5.74) is also imaginary, the product of the two factors in brackets is real, as it should be.

4.2.5.4.4. Displacements: general remarks

Even small displacements may have an important influence on the problem of propagation of order. Therefore, no structural treatments other than the introduction of formal parameters (*e.g.* Landau's theory) have been published in the literature. Most of the examples with really reliable results refer to binary systems and even these represent very crude approximations, as will be shown below. For this reason we shall restrict ourselves here to binary systems, although general formulae where displacements are included may be developed in a formal way.

Two kinds of atoms, $f_1(\mathbf{r})$ and $f_2(\mathbf{r})$, are considered. Obviously, the position of any given atom is determined by its surroundings. Their extension depends on the forces acting on the atom under consideration. These may be very weak in the case of metals (repulsive forces, the 'size effect'), but long-range effects have to be expected in ionic crystals. For the development of formulae authors have assumed that small displacements $\Delta\mathbf{r}_{\nu\nu'}$ may be assigned to the pair-correlation functions $p'_{\nu\nu'}(\mathbf{r})$ by adding a phase factor $\exp\{2\pi i\mathbf{H}\cdot\Delta\mathbf{r}_{\nu\nu'}\}$, which is then expanded in the usual way:

$$\exp\{2\pi i\mathbf{H}\cdot\Delta\mathbf{r}_{\nu\nu'}\} \simeq 1 + 2\pi i\mathbf{H}\cdot\Delta\mathbf{r}_{\nu\nu'} - 2[\pi\mathbf{H}\cdot\Delta\mathbf{r}_{\nu\nu'}]^2. \tag{4.2.5.75}$$

The displacements and correlation probabilities are separable if the change of atomic scattering factors in the angular range considered may be neglected. The formulae in use are given in the next section. As shown below, this method represents nothing other than a kind of average over certain sets of displacements. For this purpose, the correct solution of the problem has to be discussed. In the simplest model the displacements are due to next-nearest neighbours only. It is further assumed that the configurations rather than the displacements determine the position of the central atom and a general displacement of the centre of the first shell does not occur (there is no influence of a strain field). Obviously, the formal correlation function of pairs is not independent of displacements. This difficulty may be avoided either by assuming that the pair-correlation function has already been separated from the diffraction data, or by theoretical calculations of the correlation function (the mean-field method) (Moss, 1966; de Fontaine, 1972, 1973). The validity of this procedure is subject to the condition that the displacements have no influence on the correlation functions themselves.

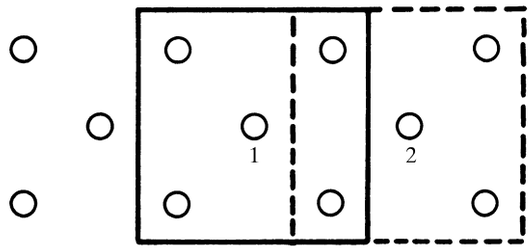


Fig. 4.2.5.1. Construction of the correlation function in the method of overlapping clusters.

The observation of a periodic average structure justifies the definition of a periodic array of origins, which normally depends on the degree of order. Local deviations of origins may be due to fluctuations in the degree of order and due to the surrounding atoms of a given site. For example, a b.c.c. lattice with eight nearest neighbours is considered. It is assumed that only these have an influence on the position of the central atom owing to different forces of the various configurations. With two kinds of atoms, there are $2^9 = 512$ possible configurations of the cluster (the central atom plus 8 neighbours). Symmetry considerations reduce this number to 28. Each is characterized by a displacement vector. Hence their *a priori* probabilities and the propagation of 28 different configurations have to be determined. Since each atom has to be considered as the centre once, this problem may be treated by introducing 28 different atomic scattering factors as determined from the displacements: $f_\nu(\mathbf{r})\exp\{2\pi i\mathbf{H}\cdot\Delta\mathbf{r}_{\nu\nu'}\}$. The diffraction problem has to be solved with the aid of the propagation of order of overlapping clusters. This is demonstrated by a two-dimensional model with four nearest neighbours (Fig. 4.2.5.1). Here the central and the neighbouring cluster (full and broken lines) overlap with two sites in, for example, the x direction. Hence only neighbouring clusters with the same overlapping pairs are admitted. These restrictions introduce severe difficulties into the problem of propagation of cluster ordering which determines the displacement field. Since it was assumed that the problem of pair correlation had been solved, the cluster probabilities may be derived by calculating

$$\alpha_\nu l(\mathbf{r}) \prod_{\mathbf{n}\neq 0} p'_{\nu\nu'}(\mathbf{r}-\mathbf{n}). \tag{4.2.5.76}$$

Only next-nearest neighbours have to be included in the product. This must be performed for the central cluster ($\mathbf{r} = 0$) *and* for the reference cluster at $\mathbf{r} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$, because all are characterized by different displacements. So far, possible displacement of the centre has not been considered; this may also be influenced by the problem of propagation of cluster ordering. These displacement factors should best be attached to the function describing the propagation of order which determines, in principle, the local fluctuations of the lattice constants (the strain field *etc.*). This may be understood by considering a binary system with a high degree of order but with atoms of different size. Large fluctuations of lattice constants are involved in the case of

exsolution of the two components because of their different lattice parameters, but they become small in the case of superstructure formation where a description in terms of antiphase domains is reasonable (with equal lattice constants). This example demonstrates the mutual dependence of ordering and displacements, which is mostly neglected in the literature.

The method of assigning phase factors to the pair-correlation function is now discussed. Pair-correlation functions average over all pairs of clusters having the same central atom. An analogous argument holds for displacements: using pair correlations for the determination of displacements means nothing other than averaging over all displacements caused by various clusters around the same central atom. There remains the general strain field due to the propagation of order, whereas actual displacements of atoms are realized by fluctuations of configurations. Since large fluctuations of this type occur in highly disordered crystals, the displacements become increasingly irrelevant. Hence the formal addition of displacement factors to the pair-correlation function does not yield much information about the structural basis of the displacements. This situation corresponds exactly to the relationship between a Patterson function and a real structure: the structure has to be found which explains the quite complicated function completely, and its unique solution is rather difficult. We state these points here because in most publications related to this subject these considerations are not taken into account adequately. Displacements usually give rise to antisymmetric contributions to diffuse reflections. As pointed out above, the influence of displacements has to be considered as phase factors which may be attached either to the structure factors or to the Fourier transforms $P'_{\nu\nu'}(\mathbf{H})$ of the correlation functions in equation (4.2.5.71). As has been mentioned in the context of equation (4.2.5.74), antisymmetric contributions will occur if acentric structure factors are involved. Apparently, this condition is met by the phase factors of displacements. In consequence, antisymmetric contributions to diffuse reflections may also originate from the displacements. This fact can also be demonstrated if the assignment of phase factors to the Fourier transforms of the correlation functions is advantageous. In this case, equations (4.2.5.69*a*,*b*) are no longer valid because the functions $p'_{\nu\nu'}(\mathbf{r})$ become complex. The most important change is the relation corresponding to (4.2.5.69):

$$\alpha_\nu p'_{\nu\nu'}(\mathbf{r}) = \alpha_{\nu'} p'^{+}_{\nu\nu'}(-\mathbf{r}) \leftrightarrow \alpha_\nu P'_{\nu\nu'}(\mathbf{H}) = \alpha_{\nu'} P'^{+}_{\nu'}(\mathbf{H}). \quad (4.2.5.77)$$

Strictly speaking, we have to replace the *a priori* probabilities α_ν by complex numbers $\alpha_\nu \exp\{2\pi i \Delta\mathbf{r}_\nu \cdot \mathbf{H}\}$, which are determined by the position of the central atom. In this way, all correlations between displacements may be included with the aid of the clusters mentioned above. To a rough approximation it may be assumed that no correlations of this kind exist. In this case, the complex factors may be assigned to the structure factors involved. Averaging over all displacements results in diffraction effects that are very similar to a static Debye–Waller factor for all structure factors. On the other hand, the thermal motion of atoms is treated similarly. Obviously, both factors affect the sharp Bragg peaks. Hence this factor can easily be determined by the average structure, which contains a Debye–Waller factor including static and thermal displacements. It should be pointed out, however, that these static displacements cause elastic diffuse scattering, which cannot be separated by inelastic neutron-scattering techniques.

A careful study of the real and imaginary parts of

$$\langle p'_{\nu\nu'}(\mathbf{r})\rangle = \langle p'_{\nu\nu'}(\mathbf{r})\rangle_{\mathrm{R}} + \langle p'_{\nu\nu'}(\mathbf{r})\rangle_{\mathrm{I}}$$

and

$$\Delta p'_{\nu\nu'}(\mathbf{r}) = \Delta p'_{\nu\nu'}(\mathbf{r})_{\mathrm{R}} + \Delta p'_{\nu\nu'}(\mathbf{r})_{\mathrm{I}}$$

and their Fourier transforms results, after some calculations, in the following relation for diffuse scattering:

$$\begin{aligned} I_{\mathrm{D}} \simeq\ & N \sum_\nu \alpha_\nu |\Delta F_\nu(\mathbf{H})|^2 \{[\langle P'_{\nu\nu}(\mathbf{H})\rangle - \Delta P'_{\nu\nu}(\mathbf{H})] * L(\mathbf{H})\} \\ & + 2N \sum_{\nu>\nu'} \alpha_\nu (\Delta F_\nu \Delta F^{+}_{\nu'})_{\mathrm{R}} \\ & \times \{[\langle P'_{\nu\nu'}(\mathbf{H})\rangle_{\mathrm{R}} - \Delta P'_{\nu\nu'}(\mathbf{H})_{\mathrm{I}}] * L(\mathbf{H})\} \\ & + 2N \sum_{\nu>\nu'} \alpha_\nu (\Delta F_\nu \Delta F^{+}_{\nu'})_{\mathrm{I}} \\ & \times \{[\langle P'_{\nu\nu'}(\mathbf{H})\rangle_{\mathrm{I}} - \Delta P'_{\nu\nu'}(\mathbf{H})_{\mathrm{R}}] * L(\mathbf{H})\}. \end{aligned} \quad (4.2.5.78)$$

It should be noted that all contributions are real. This follows from the properties of Fourier transforms of symmetric and antisymmetric functions. All $\Delta P_{\nu\nu'}(\mathbf{H})$ are antisymmetric; hence they generate antisymmetric contributions to the line profiles. In contrast to equation (4.2.5.75), the real *and* the imaginary parts of the structure factors contribute to the asymmetry of the line profiles.

4.2.5.4.5. Distortions in binary systems

In substitutional binary systems (with a primitive cell with only one sublattice) the Borie–Sparks method is widely used (Sparks & Borie, 1966; Borie & Sparks, 1971). The method is formulated in the short-range-order-parameter formalism. The diffuse scattering may be separated into two parts (*a*) owing to short-range order and (*b*) owing to static displacements.

Corresponding to the expansion (4.2.5.75), $I_{\mathrm{D}} = I_{\mathrm{sro}} + I_2 + I_3$, where I_{sro} is given by equation (4.2.5.71*b*) and the correction terms I_2 and I_3 relate to the linear and the quadratic term in (4.2.5.75). The intensity expression will be split into terms of A–A, A–B, ... pairs. More explicitly $\Delta\mathbf{r}_{\nu\nu'} = \mathbf{u}_{\mathbf{n}\nu} - \mathbf{u}_{\mathbf{n}'\nu'}$ and with the following abbreviations

$$\begin{aligned} \delta_{\mathbf{nn}'|AA} &= \mathbf{u}_{\mathbf{n}|A} - \mathbf{u}_{\mathbf{n}'|A} = x_{\mathbf{nn}'|AA}\mathbf{a} + y_{\mathbf{nn}'|AA}\mathbf{b} + z_{\mathbf{nn}'|AA}\mathbf{c} \\ \delta_{\mathbf{nn}'|AB} &= \mathbf{u}_{\mathbf{n}|A} - \mathbf{u}_{\mathbf{n}'|B} = \ldots \\ F_{\mathbf{nn}'|AA} &= [f_A^2/(f_A - f_B)^2][(c_A/c_B) + \alpha_{\mathbf{nn}'}] \\ F_{\mathbf{nn}'|BB} &= [f_B^2/(f_A - f_B)^2][(c_B/c_A) + \alpha_{\mathbf{nn}'}] \\ F_{\mathbf{nn}'|AB} &= [2f_A f_B/(f_A - f_B)^2](1 - \alpha_{\mathbf{nn}'}) = F_{\mathbf{nn}'|BA} \end{aligned}$$

one finds (where the short-hand notation is self-explanatory):

$$\begin{aligned} I_2 = 2\pi i c_A c_B (f_A - f_B)^2 \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \{ & H[F_{\mathbf{nn}'|AA}\langle x_{\mathbf{nn}'|AA}\rangle \\ & + F_{\mathbf{nn}'|BB}\langle x_{\mathbf{nn}'|BB}\rangle + F_{\mathbf{nn}'|AB}\langle x_{\mathbf{nn}'|AB}\rangle] + K[\text{`}y\text{'}] \\ & + L[\text{`}z\text{'}]\} \exp\{2\pi i \mathbf{H}\cdot(\mathbf{n} - \mathbf{n}')\}, \end{aligned} \quad (4.2.5.79)$$

$$\begin{aligned} I_3 = c_A c_B (f_A - f_B)^2 (-2\pi)^2 \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \{ & H^2[F_{\mathbf{nn}'}\langle x^2_{\mathbf{nn}'|AA}\rangle \\ & + F_{\mathbf{nn}'|BB}\langle x^2_{\mathbf{nn}'|BB}\rangle + F_{\mathbf{nn}'|AB}\langle x^2_{\mathbf{nn}'|AB}\rangle] \\ & + K^2[\text{`}y^2\text{'}] + L^2[\text{`}z^2\text{'}] \\ & + HK[F_{\mathbf{nn}'|AA}\langle (xy)_{\mathbf{nn}'|AA}\rangle + F_{\mathbf{nn}'|BB}\langle (xy)_{\mathbf{nn}'|BB}\rangle \\ & + F_{\mathbf{nn}'|AB}\langle (xy)_{\mathbf{nn}'|AB}\rangle] \\ & + KL[\text{`}(yz)\text{'}] + LH[\text{`}(zx)\text{'}]\} \\ & \times \exp\{2\pi i \mathbf{H}\cdot(\mathbf{n} - \mathbf{n}')\}. \end{aligned} \quad (4.2.5.80)$$

With further abbreviations

$$\gamma_{\mathbf{nn}'|x} = 2\pi(F_{\mathbf{nn}'|AA}\langle x_{\mathbf{nn}'|AA}\rangle + F_{\mathbf{nn}'|BB}\langle x_{\mathbf{nn}'|BB}\rangle + F_{\mathbf{nn}'|AB}\langle x_{\mathbf{nn}'|AB}\rangle)$$
$$\gamma_{\mathbf{nn}'|y} = \ldots$$
$$\gamma_{\mathbf{nn}'|z} = \ldots$$
$$\delta_{\mathbf{nn}'|x} = (-2\pi^2)(F_{\mathbf{nn}'|AA}\langle x^2_{\mathbf{nn}'|AA}\rangle + F_{\mathbf{nn}'|BB}\langle x^2_{\mathbf{nn}'|BB}\rangle + F_{\mathbf{nn}'|AB}\langle x^2_{\mathbf{nn}'|AB}\rangle)$$
$$\delta_{\mathbf{nn}'|y} = \ldots$$
$$\delta_{\mathbf{nn}'|z} = \ldots$$
$$\varepsilon_{\mathbf{nn}'|xy} = (-4\pi^2)(F_{\mathbf{nn}'|AA}\langle (xy)_{\mathbf{nn}'|AA}\rangle + F_{\mathbf{nn}'|BB}\langle (xy)_{\mathbf{nn}'|BB}\rangle + F_{\mathbf{nn}'|AB}\langle (xy)_{\mathbf{nn}'|AB}\rangle)$$
$$\varepsilon_{\mathbf{nn}'|yz} = \ldots$$
$$\varepsilon_{\mathbf{nn}'|zx} = \ldots$$
$$I_2 = c_A c_B (f_A - f_B)^2 \sum_{\mathbf{n}} \sum_{\mathbf{n}'} i(\gamma_{\mathbf{nn}'|x} + \gamma_{\mathbf{nn}'|y} + \gamma_{\mathbf{nn}'|z}) \times \exp\{2\pi i \mathbf{H}\cdot(\mathbf{n}-\mathbf{n}')\}$$
$$I_3 = c_A c_B (f_A - f_B)^2 \sum_{\mathbf{n}} \sum_{\mathbf{n}'} (\delta_{\mathbf{nn}'|x}H^2 + \delta_{\mathbf{nn}'|y}K^2 + \delta_{\mathbf{nn}'|z}L^2 + \varepsilon_{\mathbf{nn}'|xy}HK + \varepsilon_{\mathbf{nn}'|yz}KL + \varepsilon_{\mathbf{nn}'|zx}LH)\exp\{2\pi i \mathbf{H}\cdot(\mathbf{n}-\mathbf{n}')\}.$$

If the $F_{\mathbf{nn}'|AA}, \ldots$ are independent of $|\mathbf{H}|$ in a range of measurement which is better fulfilled with neutrons than with X-rays (see below), γ, δ, ε are the coefficients of the Fourier series:

$$Q_x = \sum_{\mathbf{n}} \sum_{\mathbf{n}'} i\gamma_{\mathbf{nn}'|x} \exp\{2\pi i \mathbf{H}\cdot(\mathbf{n}-\mathbf{n}')\};$$
$$Q_y = \ldots; \qquad Q_z = \ldots;$$
$$R_x = \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \delta_{\mathbf{nn}'|x} \exp\{2\pi i \mathbf{H}\cdot(\mathbf{n}-\mathbf{n}')\};$$
$$R_y = \ldots; \qquad R_z = \ldots;$$
$$S_{xy} = \sum_{\mathbf{n}} \sum_{\mathbf{n}'} \varepsilon_{\mathbf{nn}'|xy} \exp\{2\pi i \mathbf{H}\cdot(\mathbf{n}-\mathbf{n}')\};$$
$$S_{yz} = \ldots; \qquad S_{zx} = \ldots.$$

The functions Q, R, S are then periodic in reciprocal space.

The double sums over $\mathbf{n}$, $\mathbf{n}'$ may be replaced by $N\sum_{m,n,p}$ where m, n, p are the coordinates of the interatomic vectors $(\mathbf{n}-\mathbf{n}')$ and I_2 becomes

$$I_2 = -Nc_A c_B (f_A - f_B)^2 \sum_m \sum_n \sum_p (H\gamma_{lmn|x} + \ldots + \ldots) \times \sin 2\pi(Hm + Kn + Lp). \qquad (4.2.5.81)$$

The intensity is therefore modulated sinusoidally and increases with scattering angle. The modulation gives rise to an asymmetry in the intensity around a Bragg peak. Similar considerations for I_3 reveal an intensity contribution h_i^2 times a sum over cosine terms which is symmetric around the Bragg peaks ($i = 1, 2, 3$ with $h_1 = h$, $h_2 = k$, $h_3 = l$). This term shows a quite analogous influence of local static displacements and thermal movements: an increase of diffuse intensity around the Bragg peaks and a reduction of Bragg intensities, which is not discussed here. The second contribution I_2 has no analogue owing to the nonvanishing average displacement. The various diffuse intensity contributions may be separated by symmetry considerations. Once they are separated, the single coefficients may be determined by Fourier inversion. Owing to the symmetry constraints, there are relations between the displacements $\langle x \ldots \rangle$ and, in turn, between the γ and Q components. The same is true for the δ, ε and R, S components. Consequently, there are symmetry conditions for the individual contributions of the diffuse intensity which may be used to distinguish them. In general, the total diffuse intensity may be split into only a few independent terms. The single components of Q, R, S may be expressed separately by combinations of diffuse intensities that are measured in definite selected volumes in reciprocal space. Only a minimum volume must be explored in order to reveal the behaviour over the whole reciprocal space. This minimum repeat volume is different for the single components: I_{sro}, Q, R, S or combinations of them.

The Borie–Sparks method has been applied very frequently to binary and even ternary systems; some improvements have been communicated by Bardhan & Cohen (1976). The diffuse scattering of the historically important metallic compound Cu_3Au was been studied by Cowley (1950*a*,*b*) and the pair-correlation parameters could be determined. The typical fourfold splitting was found by Moss (1966) and explained in terms of atomic displacements. The same splitting has been found for many similar compounds such as Cu_3Pd (Ohshima *et al.*, 1976), Au_3Cu (Bessière *et al.*, 1983) and $Ag_{1-x}Mg_x$ (x = 0.15–0.20) (Ohshima & Harada, 1986). Similar pair-correlation functions have been determined. In order to demonstrate the disorder parameters in terms of structural models, computer programs were used (*e.g.* Gehlen & Cohen, 1965). A similar microdomain model was proposed by Hashimoto (1974, 1981, 1983, 1987). According to approximations made in the theoretical derivation, the evaluation of diffuse scattering is generally restricted to an area in reciprocal space where the influence of displacements is of the same order of magnitude as that of the pair-correlation function. The agreement between calculation and measurement is fairly good, but it should be remembered that the amount and quality of the experimental information used is low. No residual factors are so far available; these would give an idea of the reliability of the results.

The more general case of a multicomponent system with several atoms per lattice point was treated similarly by Hayakawa & Cohen (1975). Sources of error in the determination of the short-range-order coefficients are discussed by Gragg *et al.* (1973). In general, the assumption of constant $F_{\mathbf{nn}'|AA}, \ldots$ produces an incomplete separation of the order- and displacement-dependent components of diffuse scattering. By an alternative method, by separation of the form factors from the Q, R, S functions and solving a large array of linear relationships by least-squares methods, the accuracy of the separation of the various contributions is improved (Tibbals, 1975; Georgopoulos & Cohen, 1977; Wu *et al.*, 1983). The method does not work for neutron diffraction. The case of planar short-range order with corresponding diffuse intensity along rods in reciprocal space may also be treated along the Borie & Sparks method (Ohshima & Moss, 1983).

Multiwavelength methods taking advantage of the variation of the structure factor near an absorption edge (anomalous dispersion) are discussed by Cenedese *et al.* (1984). The same authors show that in some cases the neutron method allows contrast variation by using samples with different isotope substitution.

In general, X-ray and neutron methods are complementary. The neutron method is helpful in the cases of hydrides, oxides, carbides, Al–Mg distributions *etc.* In favourable cases it is possible to suppress (nuclear) Bragg scattering of neutrons when isotopes are used so that $\sum_\nu c_\nu f_\nu = 0$ for all equivalent positions. Another way to separate Bragg peaks is to record the diffuse intensity, if possible, at low $|\mathbf{H}|$ values. This can be achieved either by measurement at low angles or by using long wavelengths. For reasons of absorption the latter are the domain of neutron scattering. Bragg scattering is ruled out by exceeding the Bragg cut-off. In this way, 'diffuse' background owing to multiple Bragg scattering is avoided. Other diffuse-scattering contributions that increase with the $|\mathbf{H}|$ value [thermal diffuse scattering (TDS) and scattering due to long-range static displacements] are thus also minimized. Neutrons are preferable in cases where X-rays show only a small scattering contrast: (heavy) metal lattice distortions,

Huang scattering and so on should be measured at large values of $|\mathbf{H}|$. TDS can be separated by purely elastic neutron methods within the limits given by the energy resolution of an instrument. This technique is of particular importance at higher temperatures where TDS becomes remarkably strong. Neutron scattering is a good tool only in cases where (isotope/spin-) incoherent scattering is not too strong. In the case of magnetic materials, confusion with paramagnetic diffuse scattering could occur. This is also important when electrons are trapped by defects which themselves act as paramagnetic centres.

4.2.5.4.6. Comparison with the cluster method

In the general formula for diffuse scattering from random distributions equation (4.2.3.23*a*,*b*) may be used. Here $|\langle F(\mathbf{H})\rangle|^2$ describes the sharp Bragg maxima, while $|\Delta F(\mathbf{H})|^2 = \langle|F(\mathbf{H})|^2\rangle - |\langle F(\mathbf{H})\rangle|^2$ represents the contribution to diffuse scattering. Correlation effects can also be taken into account by using clusters of sufficient size if their distribution may be considered as random to a good approximation. The diffuse intensity is then given by

$$I_{\rm D}(\mathbf{H}) = \sum_\nu p_\nu |F_\nu(\mathbf{H})|^2 - \Big|\sum_\nu p_\nu F_\nu(\mathbf{H})\Big|^2, \qquad (4.2.5.82)$$

where $F_\nu(\mathbf{H})$ represents the difference structure factor of the νth cluster and p_ν is its *a priori* probability. Obviously equation (4.2.5.82) is of some use in only two cases. (1) The number of clusters is sufficiently small and meets the condition of a nearly random distribution. In principle, the structure may then be determined with the aid of refinement methods according to equation (4.2.5.82). Since the second term is assumed to be known from the average structure, the first term may be evaluated by introducing as many parameters as there are clusters involved. A special computer program for incoherent refinement has to be used if more than one representative cluster has to be introduced. In the case of more clusters constraints are necessary. (2) The number of clusters with similar structures is not limited. It may be assumed that their size distribution may be expressed by fluctuations using well known analytical expressions, *e.g.* Gaussians or Lorentzians. The distribution is still assumed to be random.

An early application of the cluster method was the calculation of the diffuse intensity of Guinier–Preston zones, where a single cluster is sufficient (see, *e.g.*, Gerold, 1954; Bubeck & Gerold, 1984). Unfortunately, no refinements of cluster structures have so far been published. The full theory of the cluster method was outlined by Jagodzinski & Haefner (1967).

A different approach to analysing extended defects in crystals is simply to construct analytically disorder models that include chemical substitution and atomic displacements around an atomic defect as consequence of the surrounding local distortions. This method works if the number of 'shells' of the first, second, ... neighbours is limited and if short- or long-range correlations between different clusters are neglected. After subtracting the underlying average structure and the calculation of the Fourier transform of the difference structure ΔF, the intensities can be fitted to the observed diffuse diffraction pattern and the parameters of the disorder model (occupancies and displacements) can be determined. For examples of this method see Neder *et al.* (1990) or Kaiser-Bischoff *et al.* (2005). Short-range correlations may simply be included by approximating the distribution function by summing over a finite number of defects (clusters) (Goff *et al.*, 1999).

4.2.5.5. Molecular crystals: orientational disorder

Molecular crystals show in principle disorder phenomena similar to those discussed in previous sections (*i.e.* substitutional or displacement disorder). Here we have to replace the structure factors $F_\nu(\mathbf{H})$ used in the previous sections by the molecular structure factors in their various orientations. These are usually rapidly varying functions in reciprocal space which may obscure the disorder diffuse scattering. Disorder in molecular crystals is treated by Guinier (1963), Amorós & Amorós (1968), Flack (1970), Epstein *et al.* (1982), Welberry & Siripitayananon (1986, 1987) and others.

A particular type of disorder is very common in molecular and also in ionic crystals: the centres of masses of molecules or ionic complexes form a perfect 3D lattice but their orientations are disordered. Sometimes these solids are called plastic crystals. In comparison, the liquid-crystalline state is characterized by an orientational order in the absence of long-range positional order of the centres of the molecules. A clear-cut separation is not possible in cases where translational symmetry occurs in a low dimension, *e.g.* in sheets or parallel to a few directions in crystal space. For discussion of these mesophases see Chapter 4.4.

An orientationally disordered crystal may be pictured by freezing molecules in different sites in one of several orientations. Local correlations between neighbouring molecules and correlations between position and orientation may be responsible for orientational short-range order. Thermal reorientations of the molecules are often related to an orientationally disordered crystal. Thermal vibrations of the centres of masses of the molecules, librational or rotational excitations around one or more axes of the molecules, jumps between different equilibrium positions or diffusion-like phenomena are responsible for diffuse scattering of dynamic origin. As mentioned above, the complexity of molecular structures and the associated large number of thermal modes complicate a separation from static disorder effects.

In general, high Debye–Waller factors are typical for scattering by orientationally disordered crystals. Consequently only a few Bragg reflections are observable. A large amount of structural information is stored in the diffuse background. It has to be analysed with respect to an incoherent and coherent part, elastic, quasielastic or inelastic nature, short-range correlations within one and the same molecule and between orientations of different molecules, and cross correlations between positional and orientational disorder scattering. Combined X-ray and neutron methods are therefore highly recommended.

(1) *General expressions*

On the assumption of a well ordered 3D lattice, a general expression for the scattering by an orientationally disordered crystal with one molecule per unit cell may be given. This is a very common situation. Moreover, orientational disorder is frequently related to molecules with an overall 'globular' shape and consequently to crystals of high (in particular, averaged) spherical symmetry. In the following, the relevant equations are given for this situation; these are discussed in some detail in a review article by Fouret (1979). The orientation of a molecule is characterized by a parameter ω_l, *e.g.* the set of Eulerian angles of three molecular axes with respect to the crystal axes: $l = 1, \ldots, D$ (for D possible different orientations). The equilibrium position of the centre of mass of a molecule in orientation ω_l is given by $\mathbf{r}_l$, the equilibrium position of atom k within a molecule l in orientation ω_l by $\mathbf{r}_{lk}$ and a displacement from this equilibrium position by $\mathbf{u}_{lk}$. Averaging over a long time, *i.e.* supposing that the lifetime of a discrete configuration is long compared with the period of atomic vibrations, the observed intensity may be deduced from the intensity expression corresponding to a given configuration at time t:

$$I(\mathbf{H}, t) = \sum_l \sum_{l'} F_l(\mathbf{H}, t) F_{l'}^+(\mathbf{H}, t) \times \exp\{2\pi i \mathbf{H}\cdot(\mathbf{r}_l - \mathbf{r}_{l'})\}, \qquad (4.2.5.83)$$

$$F_l(\mathbf{H}, t) = \sum_k f_k \exp\{2\pi i \mathbf{H}\cdot(\mathbf{r}_{lk} + \mathbf{u}_{lk})\}. \qquad (4.2.5.84)$$

Averaging procedures must be carried out with respect to the thermal vibrations (denoted by an overbar) and over all configurations (symbol $\langle\,\rangle$). The centre-of-mass translational vibrations and librations of the molecules are most important in this context. (Internal vibrations of the molecules are assumed to be decoupled and remain unconsidered.)

$$I(\mathbf{H},t)=\sum_l\sum_{l'}\langle\overline{F_l(\mathbf{H},t)F_{l'}^{+}(\mathbf{H},t)}\rangle \times\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_l-\mathbf{r}_{l'})\}. \quad (4.2.5.83a)$$

Thermal averaging gives (*cf.* Chapter 4.1)

$$I=\sum_l\sum_{l'}\overline{F_lF_{l'}^{+}}\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_l-\mathbf{r}_{l'})\},$$
$$\overline{F_lF_{l'}^{+}}=\sum_k\sum_{k'}f_kf_{k'}\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_{lk}-\mathbf{r}_{l'k'})\} \times\overline{\exp\{2\pi i\mathbf{H}\cdot(\mathbf{u}_{lk}-\mathbf{u}_{l'k'})\}}. \quad (4.2.5.85)$$

In the harmonic approximation $\exp\{\overline{2\pi i\mathbf{H}\cdot\Delta\mathbf{u}}\}$ is replaced by $\exp\{\frac{1}{2}|2\pi\mathbf{H}\cdot\Delta\mathbf{u}|^2\}$. This is, however, a rather crude approximation because strongly anharmonic vibrations are quite common in an orientationally disordered crystal. In this approximation $\overline{F_lF_{l'}^{+}}$ becomes

$$\overline{F_lF_{l'}^{+}}=\sum_k\sum_{k'}f_kf_{k'}\exp\{-B_k(\omega_l)\} \times\exp\{-B_{k'}(\omega_{l'})\}\exp\{D_{lk;\,l'k'}\}. \quad (4.2.5.86)$$

B_k is equal to $\frac{1}{2}\overline{(2\pi\mathbf{H}\cdot\mathbf{u}_{lk})^2}$ (the Debye–Waller factor) and depends on the specific configuration ω_l. $D_{lk;\,l'k'}=\overline{(2\pi\mathbf{H}\cdot\mathbf{u}_{lk})(2\pi\mathbf{H}\cdot\mathbf{u}_{l'k'})}$ includes all the correlations between positions, orientations and vibrations of the molecules.

Averaging over different configurations demands a knowledge of the orientational probabilities. The probability of finding molecule l in orientation ω_l is given by $p(\omega_l)$. The double probability $p(\omega_l,\omega_{l'})$ gives the probability of finding two molecules l, l' in different orientations ω_l and $\omega_{l'}$, respectively. In the absence of correlations between the orientations we have $p(\omega_l,\omega_{l'})=p(\omega_l)p(\omega_{l'})$. If correlations exist, $p(\omega_l,\omega_{l'})=p(\omega_l)p'(\omega_l|\omega_{l'})$, where $p'(\omega_l|\omega_{l'})$ defines the conditional probability that molecule l' has the orientation ω_l *if* molecule l has the orientation $\omega_{l'}$. For long distances between l and l' $p'(\omega_l|\omega_{l'})$ tends to $p(\omega_{l'})$.

The difference $\Delta(\omega_l|\omega_{l'})=p'(\omega_l|\omega_{l'})-p(\omega_{l'})$ characterizes, therefore, the degree of short-range orientational correlation. Note that this formalism corresponds fully to the $p_\mu, p_{\mu\mu'}$ used in the context of translational disorder.

The average structure factor, sometimes called averaged form factor, of the molecule is given by

$$\langle F_l\rangle=\sum_{\omega_l}p(\omega_l)F_l(\omega_l). \quad (4.2.5.87)$$

(*a*) Negligible correlations between vibrations of different molecules (Einstein model):

$$D_{lk;\,l'k'}=0 \text{ for } l\neq l'.$$

From (4.2.5.86) it follows that (the prime symbol takes the Debye–Waller factor into account)

$$\overline{\langle I\rangle}=N^2|\langle F'\rangle|^2L(\mathbf{G}) + N\Big\{\sum_k\sum_{k'}\sum_{\omega_l}p(\omega_l)f_kf_{k'}\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_{lk}-\mathbf{r}_{lk'})\} \times\exp\{D_{lk;\,lk'}\}-|\langle F'\rangle|^2\Big\} + N\sum_{\Delta l\neq0}\sum_{\omega_l}\sum_{\omega_{l'}}p(\omega_l)\Delta(\omega_l|\omega_{l'}) \times F_l'(\omega_l)F_{l'}^{+}(\omega_{l'}) \times\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_l-\mathbf{r}_{l'})\}. \quad (4.2.5.88)$$

$L(\mathbf{G})$ is the reciprocal lattice of the well defined ordered lattice. The first term describes Bragg scattering from an averaged structure. The second term governs the diffuse scattering in the absence of short-range orientational correlations. The last term takes the correlation between the orientations into account.

If rigid molecules with centre-of-mass translational displacements and negligible librations are assumed, which is a first approximation only, $|\langle F\rangle|^2$ is no longer affected by a Debye–Waller factor.

In this approximation the diffuse scattering may therefore be separated into two parts:

$$N(\langle F^2\rangle-|\langle F'\rangle|^2)=N(F^2-|\langle F\rangle|^2)+N(|\langle F\rangle|^2-|\langle F'\rangle|^2) \quad (4.2.5.89)$$

with

$$\langle F^2\rangle=\sum_{\omega_l}\sum_{\omega_{l'}}\sum_k\sum_{k'}f_k(\omega_l)f_{k'}(\omega_{l'})p(\omega_l) \times\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_{lk}-\mathbf{r}_{l'k'})\}. \quad (4.2.5.90)$$

The first term in (4.2.5.89) gives the scattering from equilibrium fluctuations in the scattering from individual molecules (diffuse scattering without correlations), the second gives the contribution from the centre-of-mass thermal vibrations of the molecules.

(*b*) If intermolecular correlations between the molecules cannot be neglected, the final intensity expression for diffuse scattering is very complicated. In many cases these correlations are caused by dynamical processes (see Chapter 4.1). A simplified treatment assumes the molecule to be a rigid body with a centre-of-mass displacement $\mathbf{u}_l$ and neglects vibrational–librational and librational–librational correlations: $D_{l;\,l'}=\overline{(2\pi\mathbf{H}\cdot\mathbf{u}_l)(2\pi\mathbf{H}\cdot\mathbf{u}_{l'})}$ $(l\neq l')$. The following expression approximately holds:

$$\overline{\langle I\rangle}=N^2|\langle F'\rangle|^2L(\mathbf{G}) + \Big\langle\sum_l\sum_{l'}F_l'(\omega_l)F_{l'}^{'+}(\omega_{l'})\exp\{D_{l;\,l'}\}\Big\rangle\{2\pi i\mathbf{H}\cdot(\mathbf{r}_l-\mathbf{r}_{l'})\} + N\Big[\sum_{\omega_l}\sum_{k,\,k'}p(\omega_l)f_kf_{k'}\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_{lk}-\mathbf{r}_{l'k'})\} \times\exp\{D_{lk;\,l'k'}\}-\sum_{\omega_l}\sum_{\omega_{l'}}\sum_k\sum_{k'}p(\omega_l)p(\omega_{l'})f_kf_{k'} \times\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_{lk}-\mathbf{r}_{l'k'})\} \times\exp\{D_{lk;\,l'k'}\}\Big]+\sum_{l\neq l'}\sum_{\omega_l}\sum_{\omega_l}p(\omega_l)\Delta(\omega_l|\omega_{l'}) \times F_l'(\omega_l)F_{l'}^{'+}(\omega_{l'}) \times\exp\{2\pi i\mathbf{H}\cdot(\mathbf{r}_l-\mathbf{r}_{l'})\}\exp\{D_{l;\,l'}\}. \quad (4.2.5.91)$$

Again, the first term describes Bragg scattering and the second corresponds to the average thermal diffuse scattering in the disordered crystal. Because just one molecule belongs to one unit cell only acoustic waves contribute to this part. To an approximation, the result for an ordered crystal may be used by replacing F by $\langle F'\rangle$ [Chapter 4.1, equation (4.1.3.4)]. The third term corresponds to random-disorder diffuse scattering. If librations are neglected this term may be replaced by $N(\langle F^2\rangle-\langle F\rangle^2)$. The

last term in (4.2.5.91) describes space correlations. Omission of $\exp\{D_{l;\,l'}\}$ or expansion to $\sim(1 + D_{l;\,l'})$ are further simplifying approximations.

In either (4.2.5.88) or (4.2.5.91) the diffuse-scattering part depends on a knowledge of the conditional probability $\Delta(\omega_l|\omega_{l'})$ and the orientational probability $p(\omega_l)$. The latter may be found, at least in principle, from the average structure factor.

(2) *Rotational structure (form) factor*

In certain cases and with simplifying assumptions, $\langle F\rangle$ [equation (4.2.5.87)] and $\langle \Delta F^2\rangle$ [equation (4.2.5.90)] may be calculated. Assuming only one molecule per unit cell and treating the molecule as a rigid body, one derives from the structure factor of an ordered crystal F_l

$$\langle F\rangle = \sum_k f_k \langle \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}_{lk}\}\rangle \tag{4.2.5.92}$$

and

$$\langle \Delta F^2\rangle = \sum_k \sum_{k'} f_k f_{k'} [\langle \exp\{2\pi i \mathbf{H}\cdot(\mathbf{r}_{lk} - \mathbf{r}_{l'k'})\}\rangle - \langle \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}_{lk}\}\rangle\langle \exp\{2\pi i \mathbf{H}\cdot\mathbf{r}_{l'k'}\}\rangle]. \tag{4.2.5.93}$$

If the molecules have random orientation in space the following expressions hold [see, *e.g.*, Dolling *et al.* (1979)]:

$$\langle F\rangle = \sum_k f_k j_0(\mathbf{H}\cdot\mathbf{r}_k) \tag{4.2.5.94}$$

$$\langle |\Delta F|^2\rangle = \sum_k \sum_{k'} f_k f_{k'} \{ j_0[\mathbf{H}\cdot(\mathbf{r}_k - \mathbf{r}_{k'})] - j_0(\mathbf{H}\cdot\mathbf{r}_k) j_0(\mathbf{H}\cdot\mathbf{r}_{k'})\}. \tag{4.2.5.95}$$

$j_0(z)$ is the zeroth order of the spherical Bessel functions and describes an atom k uniformly distributed over a shell of radius r_k.

In practice the molecules perform finite librations about the main orientation. The structure factor may then be found by the method of symmetry-adapted functions [see, *e.g.*, Press (1973), Press & Hüller (1973), Dolling *et al.* (1979), Prandl (1981, and references therein)].

$$\langle F\rangle = \sum_k f_k 4\pi \sum_\nu \sum_{\mu=-\nu}^{+\nu} i^\nu j_\nu(\mathbf{H}\cdot\mathbf{r}_k) C_{\nu\mu}^{(k)} Y_{\nu\mu}(\theta,\varphi). \tag{4.2.5.96}$$

$j_\nu(z)$ is the νth order of spherical Bessel functions, the coefficients $C_{\nu\mu}^{(k)}$ characterize the angular distribution of $\mathbf{r}_k$ and $Y(\theta,\varphi)$ are the spherical harmonics where $|\mathbf{H}|, \theta, \varphi$ denote polar coordinates of $\mathbf{H}$.

The general case of an arbitrary crystal, site and molecular symmetry and the case of several symmetrically equivalent orientationally disordered molecules per unit cell are treated by Prandl (1981); an example is given by Hohlwein *et al.* (1986). As mentioned above, cubic plastic crystals are common and therefore mostly studied up to now. The expression for $\langle F\rangle$ may then be formulated as an expansion in cubic harmonics $K_{\nu\mu}(\theta,\varphi)$:

$$\langle F\rangle = \sum_k f_k 4\pi \sum_\nu \sum_\mu i^\nu j_\nu(\mathbf{H}\cdot\mathbf{r}_k) C_{\nu\mu}^{\prime(k)} K_{\nu\mu}(\theta,\varphi), \tag{4.2.5.97}$$

where $C'_{\nu\mu}$ are modified expansion coefficients.

Taking into account isotropic centre-of-mass translational displacements, which are not correlated with the librations, we obtain

$$\langle F'\rangle = \langle F\rangle \exp\{-\tfrac{1}{6}H^2\langle U^2\rangle\}. \tag{4.2.5.98}$$

U is the mean-square translational displacement of the molecule. Correlations between translational and vibrational displacements are treated by Press *et al.* (1979).

Equivalent expressions for crystals with symmetry other than cubic may be found from the same concept of symmetry-adapted functions [tables are given by Bradley & Cracknell (1972)].

(3) *Short-range correlations*

The final terms in equations (4.2.5.88) and (4.2.5.91) concern correlations between the orientations of different molecules. Detailed evaluations need knowledge of a particular model. Examples are compounds with nitrate groups (Wong *et al.*, 1984; Lefebvre *et al.*, 1984), CBr_4 (More *et al.*, 1980, 1984) and many others (see Sherwood, 1979). The situation is even more complicated when a modulation wave with respect to the occupation of different molecular orientations is superimposed. A limiting case would be a box-like function describing a pattern of domains. Within one domain all molecules have the same orientation. This situation is common in ferroelectrics where molecules exhibit a permanent dipole moment. The modulation may occur in one or more directions in space. The observed intensity in this type of orientationally disordered crystal is characterized by a system of more or less diffuse satellite reflections. The general scattering theory of a crystal with occupational modulation waves follows the same lines as outlined in Section 4.2.3.5.1.

4.2.6. Disorder diffuse scattering from aperiodic crystals

The preceding sections of this chapter have either been related to disorder phenomena in conventional crystals defined by the presence of 3D translational symmetry, *i.e.* by a 3D lattice function, at least in the averaged sense, or to solids with crystalline order in only lower (2 or 1) dimensions where the ordering principle along 'nonperiodic' directions shows a gradual transition from long- to short-range order, in particular to a liquid-like behaviour and, finally, to an (almost) random ordering behaviour of structural units. So-called aperiodic crystals do not fit this treatment because aperiodicity denotes a different type of order, which is nonperiodic in 3D space but where translational order may be restored in higher-dimensional ($n > 3$) direct space. Three types of solids are commonly included in the class of aperiodic structures: (i) incommensurately modulated structures (IMSs), where structural parameters of any kind (coordinates, occupancies, orientations of extended molecules or subunits in a structure, atomic displacement parameters) deviate periodically from the average ('basic') structure and where the modulation period is incommensurate compared to that of the basic structure; (ii) composite structures (CSs), which consist of two (or more) intergrown incommensurate substructures with mutual interactions giving rise to mutual (incommensurate) modulations; and (iii) quasicrystals (QCs), which might simply be viewed as made up by n (≥ 2) different tiles – analogous to the elementary cell in crystals – which are arranged according to specific matching rules and which are decorated by atoms or atomic clusters. Most common are icosahedral (i-type) quasicrystals with 3D aperiodic order (3D 'tiles') and decagonal quasicrystals (d-phases) with 2D aperiodicity and one unique axis along which the QC is periodically ordered. The basic 'crystallography' of aperiodic crystals is given in Chapter 4.6 of this volume; see also the review articles by Janssen & Janner (1987) and van Smaalen (2004) (and further references therein). We only point out some of the aspects here to provide the background for a short discussion about disorder diffuse scattering in aperiodic crystals.

As outlined in Section 4.6.1, a d-dimensional (dD) ideal aperiodic crystal can be defined as a dD irrational section of an nD ('hyper'-)crystal with nD crystal symmetry. Corresponding to the section of the nD hypercrystal with the dD ($d = 1, 2, 3$) direct physical (= 'external' or 'parallel') space we have a projection of the (weighted) nD reciprocal hyperlattice onto the dD reciprocal physical space. The occurrence of Bragg reflections as a signature

of an ordered (in at least an averaged sense) aperiodic crystal form a countable dense pattern of the projected reciprocal (hyper-)lattice vectors. Disorder phenomena, *i.e.* deviations from the periodicity in higher-dimensional direct (hyper-)space, are thus related to diffuse phenomena in reciprocal (hyper-)space projected down to the reciprocal physical space. Therefore only fluctuations from the aperiodic order infer diffuse scattering in its true sense, which might be hard to discern in a dense pattern of discrete (Bragg) reflections. As a consequence, in the higher-dimensional description of aperiodic crystals the atoms must be replaced by 'hyperatoms' or 'atomic surfaces', which are extended along $(n - d)$ dimensions. If, for example, an IMS exhibits a modulation in only one direction (1D IMS), we have $n = 2$ and a 1D atomic surface (which is a continuous modulation function extended along the 1D 'internal' or 'perpendicular' subspace). The physical subspace is spanned by 1 (and +2 'unaffected') dimensions. In a second example, a decagonal QC with aperiodic order in two dimensions is described by $n = 4$ (plus the remaining coordinate along the unaffected periodic direction) and $d = 2$, *i.e.* by 2D atomic surfaces. The atomic surfaces are, as shown in Chapter 4.6, continuous n-dimensional objects for IMSs and CSs, and discrete ones in the case of QCs. In addition to the 'hyperlattice' fluctuations within the physical (parallel) subspace we have therefore an additional quality of disorder phenomena related to positions and shape and size fluctuations of the atomic surfaces within the perpendicular subspace. Therefore, one has to consider disordering effects in aperiodic crystals related to fluctuations in the (external) physical subspace as well as in the internal subspace, but one should bear in mind that the two types are often coupled. Disordering due to displacements along directions within the physical subspace are – in the context of aperiodic crystals – commonly termed 'phonon-like', in contrast to 'phason-like' disorder related to displacements parallel to directions within the internal subspace. The term phason originates from a particular type of (dynamical) fluctuations of an IMS (see below).

4.2.6.1. Incommensurately modulated structures

Quite generally, small domains with periodic or aperiodic structures give rise to broadening of reflections. Diffuse satellite scattering is due to the limited coherence length of a modulation wave within a crystal or due to short-range order of twin or nanodomains (with an 'internal' modulated structure) embedded in a periodic matrix structure. This type of diffuse scattering can be treated by the rules outlined in Section 4.2.3.5. Structural fluctuations with limited correlation lengths are, for example, responsible for diffuse scattering in 1D organic conductors (Pouget, 2004). In some cases, modulated structures are intermediate phases within a limited temperature range between a high- and low-temperature phase (*e.g.* quartz) where the structural change is driven by a dynamical instability (a soft mode). In addition to (inelastic) soft-mode diffuse scattering, dynamic fluctuations of phase and amplitude of the modulation wave give rise to diffuse intensity. In particular, the phase fluctuations, *phasons*, give rise to low-frequency scattering which might be observable as an additional diffuse contribution around condensing reflections.

4.2.6.2. Composite structures

In the case of CSs, diffuse scattering relates to interactions between the component substructures which are responsible for mutual modulations: each substructure becomes modulated with the period of the other. If one of the substructures is low-dimensional, for example chain-like structural elements embedded in tubes of a host structure, one observes diffuse planes if direct interchain correlations are mostly absent. The diffuse planes are, however, not only due to the included subsystem, but also due to corresponding modulations of the matrix structure. On the other hand, maxima superimposed on the diffuse planes reflect the influence of the matrix structure on the chain system. The thickness of the diffuse planes depends on the degree of short- or long-range order along the unique direction (*cf.* Section 4.2.5.3). This might be either a consequence of faults in the surrounding matrix which interrupt the (longitudinal) coherence of the chains (Rosshirt *et al.*, 1991) or, intrinsically, due to the degree of misfit between the periods of the mutually incommensurate structures of the host and guest structures. If this misfit becomes large, *e.g.* as a consequence of different thermal expansion coefficients of the two substructures, the modulation is only preserved within small domains and one observes a series of diffuse satellite planes (Weber *et al.*, 2000). Equivalent considerations relate to composite systems made up from stacks of planar molecules (*e.g.* van Smaalen *et al.*, 1998) or an intergrowth of layer-like substructures where the modulation is mainly due to a one-dimensional stacking along the normal to the layers. Correspondingly, quite extended diffuse streaks occur at positions of the satellites in reciprocal space.

A further discussion of the more complicated disorder diffuse scattering of higher- (than one-) dimensionally modulated IMSs and CSs and the corresponding diffuse patterns in reciprocal space is beyond the scope of this chapter. We refer to Section 4.6.3 and the references cited therein. For examples, see also Petricek *et al.* (1991).

4.2.6.3. Quasicrystals

As in the case of conventional crystals, there is no unique theory of diffuse scattering by quasicrystals. Chemical disorder, phonon- and phason-like displacive disorder, topological glass-like disorder and domain disorder exist. The term domain covers those with an aperiodic structure as well as periodic approximant domains, which are also known as approximant phases (*cf.* Section 4.6.3). Approximant phases exhibit local atomic clusters which do not differ significantly from those of related aperiodic QCs. In d-phases, disorder diffuse scattering occurs which is related to the periodic direction. Reviews of disorder diffuse scattering from quasicrystals are given, *e.g.*, by Steurer & Frey (1998) and Estermann & Steurer (1998).

Chemical disorder. Many of the quasicrystals are ternary intermetallic phases consisting of atomic clusters where two components are transition metals (TMs) with only a small difference in Z (the number of electrons). Most of them exhibit a certain amount of substitutional (chemical) disorder, in particular with respect to the distribution of the TMs. This behaviour is equally true in the approximant phases. Chemical short-range order between the TMs, if any, is largely uninvestigated because conventional X-ray diffraction and electron-microscopy methods are not sensitive enough to provide significant contrast between the TMs. If the X-ray form-factor difference Δf is small, X-ray patterns do not show diffuse scattering, whereas an analogous neutron pattern could reveal a diffuse component due to the different scattering contrast Δb of the TMs (where the b's are the neutron scattering lengths). In practice, the chemical disorder phenomena might be more complex as the compositional stability of the QCs is often rather extended and the atomic distribution in a sample is not always structurally homogeneous. Depending on the crystal-growth process, microstructures may occur with locally coexisting small QC domains with fluctuating chemical content and, moreover, coexisting QC and approximant domains. Chemical disorder might also be caused by phasonic disorder.

Phonon-type (static or dynamic) displacements, fluctuations or straining relate to the physical subspace coordinates giving rise to continuous local distortions of the atomic structure. Phason-type disorder describes, as indicated above, discontinuous atomic jumps between different sites in an aperiodic structure. Qualitatively, dominant phonon- or phason-related scattering may be separated by analysing the dependence of the diffracted inten-

sities on the components of the scattering vector in the external and the internal subspace, H_e (Q_e) and H_i (Q_i), respectively (*cf.* above and also Section 4.6.3). There are different kinds of phason-like disorder, including random phason fluctuations, phason-type modulations and phason straining, and also shape and size fluctuations of the hyperatomic surfaces. If displacing an atomic surface (hyperatom) parallel to an internal space component by any kind of phason fluctuation which is equivalent to an infinitesimal rotation of the external (physical) space, it might happen that the hyperatom no longer intersects the physical space. Then an empty site is created, which is compensated by a (real) atom 'appearing' at a different position in physical space. In addition, there might even be a change of atomic species as a hyperatom may be chemically different at different sites of the atomic surface. Randomly distributed phason strains are responsible for Bragg-peak broadening and Huang-type diffuse scattering close to Bragg peaks. There are well developed theories based on the elastic theory of icosahedral (*e.g.* Jaric & Nelson, 1988) or decagonal (Lei *et al.*, 1999) quasicrystals that also include phonon–phason coupling. An example of the quantitative analysis of diffuse scattering by an i-phase (Al–Pd–Mn) is given by de Boissieu *et al.* (1995) and by a d-phase (Al–Ni–Fe) by Weidner *et al.* (2004). Dislocations in quasicrystals have partly phonon- and partly phason-like character; a discussion of the specific dislocation-related diffuse scattering will not be given here (*cf.* Section 4.6.3). Arcs and rings of localized diffuse scattering are observed in various i-phases and could be modelled in terms of some short-range 'glass-like' ordering of icosahedral clusters (Goldman *et al.*, 1988; Gibbons & Kelton, 1993). Phason-related diffuse scattering phenomena are discussed in more detail in Section 4.6.3. Depending on the exact stoichiometry and the growth conditions, d-phases very often show intergrown domain structures where the internal atomic structure of an individual domain varies between a (periodic) approximant structure, more or less strained aperiodic domains, or transient aperiodic variants (Frey & Weidner, 2002). Apart from finite size effects of domains of any kind which cause peak broadening, there are complex diffraction patterns of satellite reflections, diffuse maxima and diffuse streaking in d-Al–Ni–Co and other d-phases (Weidner *et al.*, 2001). In various d-phases one also observes, in addition to the Bragg layers, prominent diffuse layers perpendicular to the unique 'periodic' axis. They correspond to n-fold (n = 2, 4, 8) superperiods along this direction. In different d-phases there is a gradual change from almost completely diffuse planes to layers of superstructure reflections (satellites). The picture of a kind of stacking disorder of aperiodic layers does not match such observations. An explanation is rather due to 1D columns of atomic icosahedral clusters along the unique direction (Steurer & Frey, 1998). The temperature behaviour of these diffuse layers was studied by *in situ* neutron diffraction (Frey & Weidner, 2003), which shed some light on the complicated order/disorder phase transitions in d-Al–Ni–Co.

4.2.7. Computer simulations and modelling

4.2.7.1. Introduction

The various analytical expressions given in Section 4.2.5 are mostly rather complex, so their application is often restricted to relatively simple systems. Even in these cases analytical solutions are often not available. For larger displacements the approximations that use an expansion of the exponentials [*e.g.* equation (4.2.3.27)] are no longer valid. Hence there is a need for alternative approaches to tackling more complicated disorder models. In the past, optical transforms (*e.g.* Harburn *et al.*, 1974, 1975) and the videographic method (Rahman, 1993) were developed for this purpose. In the first method, a 2D mask with holes with sizes that represent the scattering power of the atoms is generated, which is then subjected to coherent light (from a laser) to produce a diffraction image. Problems arise for strong scatterers requiring very large holes. These problems were overcome in the second method, where the mask is replaced by a computer image with intensities proportional to the scattering power for each pixel. With the advent of more and more powerful computers these methods are now replaced by complete computer simulations, both to set up the disorder model and to calculate the diffraction pattern.

4.2.7.2. Simulation programs

Having established a disordered crystal with the types and positions of all atoms involved (a configuration), *e.g.* by using one of the methods described below, computer programs employing fast-Fourier-transform techniques can be used to calculate the diffraction pattern, which may be compared with the observation. It has to be borne in mind, however, that there is still a large gap between a real crystal with its ~10^{23} atoms and the one simulated by the computer with only several thousand atoms. This means that very long range correlations can not be included and have to be treated in an average manner. Furthermore, the limited size of the simulated crystal leads to termination effects, giving rise to considerable noise in the calculated diffraction pattern. Butler & Welberry (1992) have introduced a technique to avoid this problem in their program *DIFFUSE* by dividing the simulated crystal into smaller 'lots'. For each lot the intensity is calculated and then the intensities of all lots are summed up incoherently, which finally results in a smooth intensity distribution. Note, however, that in this case long-range correlations are restricted to even smaller values. To overcome this problem Boysen (1997) has proposed a method for suppressing the subsidiary maxima by multiplying the scattering density of the model crystal by a suitably designed weighting function simulating the effect of the instrumental resolution function.

A very versatile computer program, *DISCUS*, which allows not only the calculation of the scattering intensities but also allows the model structures to be built up in various ways, has been designed by Proffen & Neder (1999). It contains modules for reverse Monte Carlo (RMC, see below) simulation, the calculation of powder patterns, RMC-type refinement of pair-distribution functions (PDFs, see below) and many other useful tools for analysing disorder diffuse scattering.

Other computer programs have been developed to calculate diffuse scattering, some for specific tasks, such as *SERENA* (Micu & Smith, 1995), which uses a collection of atomic configurations calculated from a molecular dynamics simulation of molecular crystals.

4.2.7.3. Modelling procedures

Several well established methods can be used to create the simulated crystal on the computer. With such a crystal at hand, it is possible to calculate various thermodynamical properties and study the effect of specific parameters of the underlying model. By Fourier transformation, one obtains the diffraction pattern of the total scattering, *i.e.* the diffuse intensities and the Bragg peaks. The latter may lead to difficulties due to the termination effects mentioned in Section 4.2.7.2. These may be circumvented by excluding regions around the Bragg peaks (note, however, that in this case valuable information about the disorder may be lost), by subtracting the average structure or by using the approximation of Boysen (1997) mentioned above. A major advantage of such modelling procedures is that realistic physical models are introduced at the beginning, providing further insight into the pair interactions of the system, which can only be obtained *a posteriori* from the correlation parameters or fluctuation wave amplitudes derived from one of the methods described in Section 4.2.5.

4.2.7.3.1. Molecular dynamics

Molecular dynamics (MD) techniques have been developed to study the dynamics of a system. They may also be used to study static disorder problems (by taking time averages or snapshots), but they are particularly useful in the case of dynamic disorder, *e.g.* diffusing atoms in superionic conductors. The principle is to set up a certain configuration of atoms with assumed interatomic potentials $\Phi_{ij}(r_{ij})$ and subject them to Newton's equation of motion,

$$F_i(t) = m_i \frac{\mathrm{d}^2 r_i(t)}{\mathrm{d}t^2}, \qquad (4.2.7.1)$$

where the force $F_i(t)$ is calculated from the gradient of Φ. The equations are solved approximately by replacing the differential $\mathrm{d}t$ by a small but finite time step Δt to find new positions $r_i(t + \Delta t)$. This is repeated until an equilibrium configuration is found. MD techniques are quite useful if only short-range interactions are effective, even allowing the transfer of potential parameters between different systems, but are less reliable in the presence of significant long-range interactions.

4.2.7.3.2. Monte Carlo calculations

Monte Carlo (MC) methods appear to be more suited to the study of static disorder and many examples of their application can be found in the literature. Different variants allowing simulations and refinements have been applied:

(1) *Direct MC simulation*. In this method, introduced by Metropolis *et al.* (1953), a starting configuration is again set up in accordance with the known average structure and other crystal, chemical and physical information if available, and an appropriate interaction potential is chosen. The choice of this potential can be a quite delicate task. On the one hand, it should be as simple as possible to allow as large a simulation box size as can reasonably be handled by the computer's capacity. On the other hand, it must be detailed enough to include all the relevant interaction parameters of the system. Frequently used interaction potentials are a pseudo spin Ising Hamiltonian, where the 'spins' can either be binary (*e.g.* atom types or molecular orientations) or continuous (displacements), and harmonic springs for the displacements.

Having set up the starting configuration and defined the Hamiltonian, one proceeds by choosing a specific site at random and changing its parameters (occupancies and positions) by a random amount (a 'move'). The energy of this new configuration is calculated and compared with the old one. If the difference ΔE is negative, the move is accepted. If it is positive it is accepted with a probability

$$P = \exp\{-\Delta E/k_\mathrm{B}T\}/[1 + \exp\{-\Delta E/k_\mathrm{B}T\}], \qquad (4.2.7.2)$$

where T is a temperature and k_B is Boltzmann's constant. Then another site is chosen at random and the process is repeated again and again until an equilibrium configuration is found, *i.e.* until the energies fluctuate around some average value. After this, the corresponding intensity distribution is calculated and compared with the observed one. The parameters of the Hamiltonian may then be modified to improve the agreement between the calculated and observed diffraction patterns. This rather cumbersome trial-and-error method may be used to study the influence of the various parameters of the model. Further details of this method together with some illustrative examples may be found in Welberry & Butler (1994).

In an attempt to automate the adjustment of the model parameters, Welberry *et al.* (1998) have used a least-squares procedure to minimize

$$\chi^2 = \sum w(\Delta I)^2, \qquad (4.2.7.3)$$

where ΔI are the differences between the observed and calculated intensities and w are the appropriate weights. The problem with this approach is that the necessary differentials $\delta\Delta I/\delta p_i$ must be calculated numerically by performing full MC simulations with parameters p_i and $p_i + \delta p_i$ at each iteration step, which presents a formidable task even for the fastest modern computers.

(2) *Reverse MC calculations*. Application of the direct MC method may be very time consuming, as many MC simulations are necessary to arrive at a final configuration. To overcome this problem, the so-called reverse Monte Carlo (RMC) method has been developed, originally for liquids and glasses (McGreevy & Pusztai, 1988) and later for disordered single crystals (Nield *et al.*, 1995).

RMC is a model-free approach, *i.e.* without the need to define a proper interaction Hamiltonian. Otherwise it proceeds in a very similar way to direct MC analysis. A starting configuration is set up and random 'moves' of the atoms are carried out. The only difference is that acceptance or non-acceptance of a move is based on the agreement between observed and calculated diffraction intensities, *i.e.* equation (4.2.7.2) is replaced by

$$P = \exp\{-\Delta\chi^2/2\}, \qquad (4.2.7.4)$$

where χ^2 is defined in equation (4.2.7.3) and $\Delta\chi^2 = \chi^2_\mathrm{new} - \chi^2_\mathrm{old}$. Usually only a single MC run is necessary to arrive at a configuration with a diffraction pattern consistent with the observation. A drawback of the RMC technique is that usually only one configuration is found satisfying the observed diffraction pattern. In fact, it is possible that different configurations may produce the same or very similar intensity distributions (see *e.g.* Welberry & Butler, 1994). In this context, it has to be borne in mind that the diffuse scattering contains information about two-body correlations only, while the physical reason for a particular disordered structure may also be influenced by three- and many-body correlations. Such many-body correlations can easily be incorporated in the direct MC method. Moreover, RMC is susceptible to fitting artefacts (noise) in the diffraction pattern. To avoid unreasonable atomic configurations, restrictions such as a limiting nearest-neighbour approach can be built in. Weak constraints, *e.g.* ranges of interatomic distances or bond angles (*i.e.* three-body correlations) can be introduced by adding further terms in the definition of χ^2 [equation (4.2.7.3)]. In the same way, multiple data sets (*e.g.* neutron and X-ray data, EXAFS measurements *etc.*) may be incorporated. For more details and critical reviews of this method see *e.g.* McGreevy (2001) and Welberry & Proffen (1998). Many applications of this method can be found in the literature, mainly for powder diffraction, but also for single-crystal data.

(3) *Simulated annealing and evolutionary algorithms*. A general problem with MC methods is that they may easily converge to some local minima without having found the global one. One way to reduce such risks is to start with a high probability in (4.2.7.2) by using a large 'temperature' T, *i.e.* initially allowing many 'false' moves before gradually reducing T during the course of the simulation cycles. This is called 'simulated annealing', although it has nothing to do with real annealing. In the RMC method one may introduce a weighting parameter similar to T in (4.2.7.2).

Another effective way to find the global minimum and also to accelerate the optimization of the energy parameters of a given MC model is by using the principles of evolutionary theory: selection, recombination and mutation. Two such evolutionary (or genetic) algorithms have been proposed: the differential evolution (DE) algorithm (Weber & Bürgi, 2002) and the cooperative evolution (CE) algorithm (Weber, 2005). A single parameter of the model is called a gene and a set of genes is called a chromosome, p being the genotype of an individual. First a population consisting of several individuals is built up and the corresponding diffraction patterns are calculated and compared

with the observation. The fitness of each individual is quantified by χ^2. Children are then created by choosing one parent individual and calculating the second one from three randomly chosen individuals according to

$$p'_c = p_c + f_m(p_a - p_b), \qquad (4.2.7.5)$$

where f_m governs the mutations. The chromosome of the child is obtained by combining the genes of the two parents governed by a crossover (or recombination) constant f_r. If its fitness is higher than that of the parent, it replaces it. This procedure is repeated until some convergence criterion is reached. This DE method is still rather time-consuming on the computer. Therefore, the CE technique was introduced, where only one crystal is built up during the refinement. Here a large population is created spanning a large but reasonable parameter space. Individuals are selected at random to decide upon acceptance or rejection of an MC move *via* their own energy criteria. Then χ^2 is calculated and according to its positive or negative change a gratification or penalty weight is given to that individual. It may live as long as this weight is positive, otherwise it is replaced by a new individual calculated from (4.2.7.5). This way, useful individuals live longer to act on the same crystal, while unsuccessful ones are eliminated early. Note that the recombination operation is not used in this technique.

(4) *The PDF method*. The pair-distribution function (PDF) has long been used for the analysis of liquids, glasses and amorphous substances (Warren, 1969), but has recently regained considerable interest for the analysis of crystalline substances as well (Egami, 1990; Billinge & Egami, 1993; Egami, 2004). The PDF is obtained by a Fourier transformation of the total (Bragg plus diffuse) scattering in a powder pattern,

$$\rho_0 G(r) = \rho_0 + \tfrac{1}{2}\int H[S(H) - 1]\sin(2\pi H)\ \mathrm{d}H. \qquad (4.2.7.6)$$

This is nothing other than the van Hove correlation function (4.2.2.2) or the related Patterson function (4.2.2.5) averaged spherically and taken at $t = 0$ (a snapshot) and is given by

$$G^{\mathrm{PDF}}(r) = 4\pi r[\rho(r) - \rho_0] = \left(\sum c_i b_i\right)^{-2} 4\pi r \rho_0 G(r), \qquad (4.2.7.7)$$

where ρ_0 is the average number density and

$$\rho(r) = (1/N)\sum (b_i b_{i'}/\langle b\rangle^2)\delta(r - r_{ii'}). \qquad (4.2.7.8)$$

The δ functions are then convoluted with a normalized Gaussian to account for (harmonic) thermal motion. Parameter refinement proceeds in a similar way to the RMC technique. First a model is built, initially within just one unit cell and with periodic boundary conditions, then its PDF is calculated, compared with the observed one and further improved using, for example, MC simulated annealing. The model is then enlarged to include longer-range correlations. Owing to the small size of the models, this technique is much faster than the RMC method. It is essential, however, that data are measured up to very high H values to minimize truncation errors in (4.2.7.6).

4.2.7.3.3. General remarks

All of the different modelling techniques mentioned in this section have their specific merits and limitations and have contributed much to our understanding of disorder in crystalline materials following the interpretation of the corresponding diffuse scattering. It should be borne in mind, however, that application of these methods is still far from being routine work and it requires a lot of intuition to ensure that the final model is physically and chemically reasonable. In particular, it must always be ensured that the average structure remains consistent with that derived from the Bragg reflections alone. This may be done by keeping the Bragg reflections, *i.e.* by analysing the total scattering, or by designing special algorithms, *e.g.* by swapping two atoms at the same time (Proffen & Welberry, 1997). Moreover, possible traps and corrections like local minima, termination errors, instrumental resolution, statistical noise, inelasticity *etc*. must be carefully considered. All this means that the analytical methods outlined in Section 4.2.5 keep their value and should be preferred wherever possible.

The newly emerging technique of using full quantum-mechanical *ab initio* calculations for structure predictions along with MD simulations may also be applied to disordered systems. The limited currently available computer power, however, restricts this possibility to rather simple systems and small simulation box sizes, but, with the expected further increase of computer capacities, this may open up new perspectives for the future.

4.2.8. Experimental techniques and data evaluation

Single-crystal and powder diffractometry are used in diffuse scattering work. Conventional and more sophisticated special techniques and instruments are now available at synchrotron facilities and modern neutron reactor and spallation sources. The full merit of the dedicated machines may be assessed by inspecting the corresponding handbooks, which are available upon request from the facilities. In the following, some common important aspects that should be considered when planning and performing a diffuse-scattering experiment are summarized and a short overview of the techniques is given. Methodological aspects of diffuse scattering at low angles, *i.e.* small-angle-scattering techniques, and high-resolution single-crystal diffractometers are excluded. Instruments of the latter type are used when diffuse intensities beneath Bragg reflections or reflection profiles and tails must be analysed to study long-range distortion fields around single defects or small defect aggregates. In the case of small defect concentrations, the crystal structure remains almost perfect and the dynamical theory of diffraction is more appropriate. This topic is beyond the scope of this chapter.

4.2.8.1. Single-crystal techniques

In general, diffuse scattering is weak in comparison with Bragg scattering, and is anisotropically and inhomogeneously distributed in reciprocal space. The origin may be a static phenomenon or a dynamic process, giving rise to elastic or inelastic (quasi-elastic) diffuse scattering, respectively. If the disorder problem relates to more than one structural element, different parts of the diffuse scattering may show different behaviour in reciprocal space and/or on an energy scale. Therefore, before starting an experiment, some principal aspects should be considered: Is there need for X-ray and/or neutron methods? What is the optimum wavelength or energy (band), or does a 'white' technique offer advantages? Can focusing techniques be used without too strong a loss of resolution and what are the best scanning procedures? How can the background be minimized? Has the detector a low intrinsic noise and a high dynamic range?

On undertaking an investigation of a disorder problem by an analysis of the diffuse scattering, an overall picture should first be recorded by X-ray diffraction. Several sections through reciprocal space help to define the problem. For this purpose 'old-fashioned' film methods may be used, where the classical film is now commonly replaced by an imaging plate (IP) or a charge-coupled device (CCD) camera (see below). Clearly, short crystal-to-detector distances provide larger sections and avoid long exposure times, but suffer from spatial resolution. Distorted sections through the reciprocal lattice, such as produced by the Weissenberg method, may be transformed into a form suitable for easy interpretation (Welberry, 1983). The transformation of diffuse data measured using an IP or CCD requires suitable

software for the specific type of detector and is not always routinely available (Estermann & Steurer, 1998).

Standing-crystal techniques in combination with monochromatic radiation, usually called monochromatic Laue techniques (see, *e.g.*, Flack, 1970), save exposure time, which is particularly interesting for 'in-house' laboratory diffuse X-ray work. The Noromosic technique (Jagodzinski & Korekawa, 1973) is characterized by a convergent monochromatic beam which simulates an oscillation photograph over a small angular range. Heavily overexposed images, with respect to Bragg scattering, allow for sampling of diffuse intensity if a crystal is oriented in such a way that there is a well defined section between the Ewald sphere and the diffuse phenomenon under consideration. By combining single Noromosic photographs, Weissenberg patterns can be simulated. This relatively tedious method is often unavoidable because the heavily overexposed Bragg peaks obscure weak diffuse phenomena. Furthermore, standing pictures at distinct crystal settings in comparison with conventional continuous recording are frequently sufficient in diffuse scattering work and save time. Long-exposure Weissenberg photographs are therefore not equivalent to a smaller set of standing photographs. In this context it should be mentioned that a layer-line screen has not only the simple function of a selecting diaphragm, but the gap width determines the resolution volume within which diffuse intensity is collected (Welberry, 1983). For further discussion of questions of resolution see below. A comparison of Weissenberg and diffractometer methods for the measurement of diffuse scattering is given by Welberry & Glazer (1985). It should be pointed out, however, that diffractometer methods at synchrotron sources become more widely used if only very small (micrometre-sized) single-crystal specimens have to be used to study a disorder problem.

The basic arguments for using neutron-diffraction methods were given in Section 4.2.2.2: (i) the different interactions of X-rays and neutrons with matter; (ii) the lower absorption of neutrons, in particular when using longer (> 0.15 nm) wavelengths; and (iii) the matching of the energy of thermal neutrons with that of the phonons that contribute to the TDS background, and, in consequence, to separate it by a 'purely' elastic measurement. A comparative consideration of synchrotron- and neutron-related diffuse work on disordered alloys is given by Schweika (1998). Specific aspects of neutron diffraction and instruments are discussed at the end of this section.

Intensities: As mentioned above, diffuse intensities are usually weaker by several orders of magnitude than Bragg data. Therefore intense radiation sources are needed. Even a modern X-ray tube is a stronger source, defined by the flux density from the anode (number of photons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$), than modern neutron sources. For this reason most experimental work which can be performed with X-rays should be. In home laboratories the intense characteristic spectrum of an X-ray tube is commonly used. At synchrotron storage rings any wavelength from a certain range can be selected. The extremely high brilliance (number of quanta cm^{-2}, sr^{-1}, s^{-1} and wavelength interval) of modern synchrotron sources is, however, unnecessary in the case of slowly varying diffuse phenomena. In these cases, an experimental setup at a laboratory rotating anode is competitive and often even superior if specimens with sufficient size are available. Various aspects of diffuse X-ray work at a synchrotron facility are discussed by Matsubara & Georgopoulos (1985), Oshima & Harada (1986) and Ohshima *et al.* (1986). Diffuse neutron-diffraction work can only be performed on a high-flux reactor or on a powerful spallation source. Highly efficient monochromator systems are needed when using a crystal diffractometer. Time-of-flight (TOF) neutron diffractometers at pulsed (spallation) neutron sources are equivalent to conventional diffractometers at reactors (Windsor, 1982). The merits of diffuse neutron work at pulsed sources have been discussed by Nield & Keen (2001).

Wavelength: The choice of an optimum wavelength is important with respect to the problem to be solved. For example, point defects cause diffuse scattering to fall off with increasing scattering vector; short-range ordering between clusters causes broad peaks corresponding to large d spacings; lattice-relaxation processes induce a broadening of the interferences; static modulation waves with long periods give rise to satellite scattering close to Bragg peaks. In all these cases, a long wavelength is preferable due to the higher resolution. The use of a long wavelength is also profitable when the main diffuse contributions can be recorded within an Ewald sphere as small as the Bragg cutoff of the sample at $H \simeq 1/(2d_{\max})$. 'Contamination' by Bragg scattering can thus be avoided. This is also advantageous from a different point of view: because the contribution of thermal diffuse scattering increases with increasing scattering vector $\mathbf{H}$, the relative amount of this component becomes negligibly small within the first reciprocal cell. However, one has to take care with the absorption of long-wavelength neutrons.

On the other hand, short wavelengths are needed where atomic displacements play the dominant role. If diffuse peaks in large portions of the reciprocal space, or diffuse streaks or planes, have to be recorded up to high values of the scattering vector in order to decide between different structural disorder models, hard X-rays or hot neutrons are needed. For example, high-energy X-rays (65 keV) provided at a synchrotron source were used by Welberry *et al.* (2003) to study diffuse diffraction from ceramic materials and allowed studies with better d resolution.

The λ^3 dependence of the scattered intensity, in the framework of the kinematical theory, is a crucial point for exposure or data-acquisition times. Moreover, the accuracy with which an experiment can be carried out suffers from a short wavelength: generally, momentum as well as energy resolution are lower. For a quantitative estimate detailed considerations of resolution in reciprocal space (and energy) are needed.

A specific wavelength aspect concerns the method of (X-ray) anomalous dispersion, which may also be used in diffuse-scattering work. It allows the contrast and identification of certain elements in a disordered structure. Even small concentrations of impurity atoms or defects as low as 10^{-6} can be determined by this method if the impurity atoms are located at specific sites, *e.g.* in domain boundaries, or if certain other defect structures exist with characteristic diffuse scattering in reciprocal space. The (weak) diffuse scattering can then be contrasted by tuning the wavelength across an absorption edge of the particular atomic species. To avoid strong fluorescence background scattering such an experiment is usually performed at different wavelengths (energies) at the 'low absorption side' close to an edge. The merit of this method was demonstrated by Berthold & Jagodzinski (1990), who analysed diffuse streaks due to boundaries between lamellar domains in an albite feldspar. Similar element contrast can be achieved in neutron diffuse work if using specimens with different isotopes, *e.g.* H–D exchange.

Monochromacy: In classical crystal diffractometry, monochromatic radiation is used in order to eliminate broadening effects due to the wavelength distribution. Focusing monochromators and other focusing devices (guides, mirrors) help to overcome the lack of luminosity. A focusing technique is very helpful for deciding between geometrical broadening and 'true' diffuseness. In a method that is used with some success, the sample is placed in a monochromatic divergent beam with its selected axis lying in the scattering plane of the monochromator (Jagodzinski, 1968). The specimen is fully embedded in the incident beam, which is focused onto a 2D detector. Using this procedure the influence of the sample size is suppressed in one dimension. In white-beam (neutron) time-of-flight diffractometry the time resolution is the counterpart to the wavelength resolution. This is discussed in some detail in the textbook of Keen & Nield (2004).

Detectors: Valuable developments with a view to diffuse-scattering work are multidetectors (see, *e.g.*, Haubold, 1975) and position-sensitive detectors for X-rays (Arndt, 1986*a*,*b*) and neutrons (Convert & Forsyth, 1983). A classical '2D area detector' is the photographic film – nowadays of less importance – which has to be 'read out' by microdensitometer scanning. Important progress in recording diffuse X-ray data has been made by the availability of multiwire detectors, IPs and CCD cameras (Estermann & Steurer, 1998). For 2D position-sensitive (proportional) counters problems may arise from inhomogeneities of the wire array as well as the limited dynamic range when a Bragg reflection is accidentally recorded. IP systems have the major advantage of a larger dynamic range, 10^5–10^6, compared to 10^2–10^3 for an X-ray film. IPs can also be used in diffuse experiments carried out with hard (65 keV) X-rays (Welberry *et al.*, 2003). IPs are also available for neutron work, where the necessary transformation of the detected neutrons into light signals is provided by special neutron-absorbing converter foils (Niimura *et al.*, 2003). The problem of an intrinsic sensitivity to gamma radiation may be overcome by protection with a thin sheet of lead. IPs allow data collection either in plane geometry or in simple rotation or Weissenberg geometry, both in combination with low- and high-temperature devices. CCD detectors are well suited for X-ray diffuse scattering, and when used in combination with converter foils they can also be used for the detection of neutrons. A basic prerequisite is a low intrinsic noise, which can be achieved by cooling the CCD with liquid nitrogen. With these detectors, extended diffuse data sets can be collected by rotating the sample in distinct narrow steps around a spindle axis over several χ settings and subsequent oscillations over a small angular range. Thus large parts of the reciprocal space can be recorded. An example is given by Campbell *et al.* (2004), who studied subtle defect structures in the microporous framework material mordenite. Linear position-sensitive detectors are mainly used in powder work, but can also be used for recording diffuse scattering by single crystals. By combining a linear position-sensitive detector and the TOF method, a whole area in reciprocal space is accessible simultaneously (Niimura *et al.*, 1982; Niimura, 1986).

Diffuse data recorded with IPs or CCDs are commonly evaluated by commercial software supplied by the manufacturer. These software packages include corrections for extracted Bragg data, but no special software tools for treating extended (diffuse) data are commonly available. One particular problem relates to the background definition for diffuse IP data. Even after subtraction of electronic noise, there remains considerable uncertainty about the amount of true background scattering from the sample as recorded by an IP scanner. The definitions of errors and error maps remain doubtful as long as the true conversion rate between the captured neutron or photon *versus* the recorded optical signal is unknown.

Absorption: Special attention must be paid to absorption phenomena, in particular when (in the X-ray case) an absorption edge of an element of the sample is close to the wavelength. Then strong fluorescence scattering may completely obscure weak diffuse-scattering phenomena. In comparison with X-rays, the generally lower absorption coefficients of neutrons make absolute measurements easier. This also allows the use of larger sample volumes, which is not true in the X-ray case. Moreover, the question of sample environment is less serious in the neutron case than in the X-ray case. However, the availability of hard X-rays at a synchrotron source (Butler *et al.*, 2000; Welberry *et al.*, 2003) makes the X-ray absorption problem less serious: irregularly shaped specimens without special surface treatment could be used. There is also no need for complicated absorption corrections and the separation of fluorescence background is rather simple.

Extinction: An extinction problem does not generally exist in diffuse-scattering work.

Background: An essential prerequisite for a diffuse-scattering experiment is the careful suppression of background scattering. Incoherent X-ray scattering by a sample gives continuous blackening in the case of fluorescence, or scattering at high 2θ angles owing to Compton scattering or 'incoherent' inelastic effects. Protecting the image plate with a thin Al or Ni foil is of some help against fluorescence, but also attenuates the diffuse intensity. Obviously, energy-dispersive counter methods are highly efficient in this case (see below). Air scattering produces a background at low 2θ angles which may easily be avoided by special slit systems and evacuation of the camera.

In X-ray or neutron diffractometer measurements, incoherent and multiple scattering contribute to a background which varies only slowly with 2θ and can be subtracted by linear interpolation or fitting a smooth curve, or can even be calculated quantitatively and then subtracted. In neutron diffraction there are rare cases when monoisotopic and 'zero-nuclear-spin' samples are available and, consequently, the corresponding incoherent scattering part vanishes completely. In some cases, a separation of coherent and incoherent neutron scattering is possible by polarization analysis (Gerlach *et al.*, 1982). An 'empty' scan can take care of instrumental background contributions. Evacuation or controlled-atmosphere studies need a chamber, which may give rise to spurious scattering. This can be avoided if no part of the vacuum chamber is hit by the primary beam. The problem is less serious in neutron work. The mounting of the specimen, *e.g.*, on a silica fibre with cement, poorly aligned collimators and beam catchers are further sources. Sometimes a specimen has to be enclosed in a capillary, which will always be hit by the incident beam. Careful and tedious experimental work is necessary in the case of low- and high-temperature (or -pressure) investigations, which have to be carried out in many disorder problems. While the experimental situation is again less serious in neutron scattering, there are problems with scattering from walls and containers in X-ray work. Because TDS dominates at high temperatures and in the presence of a static disorder problem, a quantitative separation can rarely be carried out in the case of high experimental background. Calculation and subtraction of the TDS is possible in principle, but difficult in practice. If the disorder problem in which one is interested in does not change with temperature, a low-temperature experiment can be carried out. Another way to get rid off TDS, at least partly, is by using a neutron diffractometer with an additional analyser set to zero-energy transfer.

Resolution: A quantitative analysis of diffuse-scattering data is essential for reaching a definite decision about a disorder model, but – in many cases – it is cumbersome. By comparing the calculated and corrected experimental data the magnitudes of the parameters of the structural disorder model may be derived. A careful analysis of the data requires, therefore, after separation of the background (see above), corrections for polarization (in the X-ray case), absorption (in conventional X-ray work) and resolution. Detailed considerations of instrumental resolution are necessary; the resolution depends, in addition to other factors, on the scattering angle and implies intensity corrections analogous to the Lorentz factor used in structure analysis from sharp Bragg reflections.

Resolution is conveniently described by the function $R(\mathbf{H}-\mathbf{H}_0)$, which is defined as the probability of detecting a photon or neutron with momentum transfer $h\mathbf{H} = h(\mathbf{k}-\mathbf{k}_0)$ when the instrument is set to measure $\mathbf{H}_0$. This function R depends on the instrumental parameters (such as the collimations, the mosaic spread of monochromator and the scattering angle) and the spectral width of the source. Fig. 4.2.8.1 shows a schematic sketch of a diffractometer setting. Detailed considerations of resolution volume in X-ray and neutron diffractometry are given by Sparks & Borie (1966) and by Cooper & Nathans (1968*a*,*b*), respectively. If a triple-axis (neutron) instrument is used, for example in a purely elastic configuration, the set of instrumental parameters includes the mosaicity of the analyser

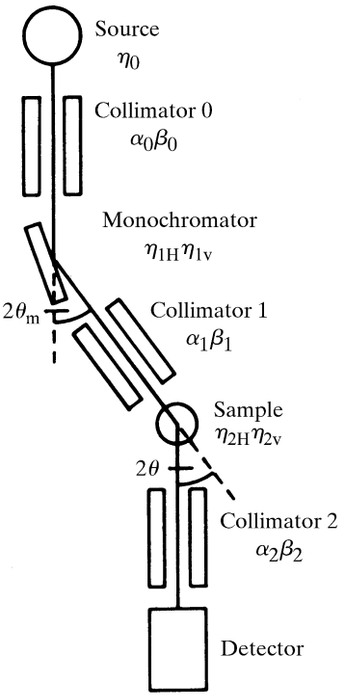


Fig. 4.2.8.1. Schematic sketch of a diffractometer setting.

and the collimations between the analyser and detector. In general, Gaussians are assumed to parameterize the mosaic distributions and transmission functions. Sophisticated resolution-correction programs are usually provided at any standard instrument for carrying out experiments at synchrotron and neutron facilities.

The general assumption of Gaussians is not too serious in the X-ray case (Iizumi, 1973). Restrictions are due to absorption, which makes the profiles asymmetric. Box-like functions are considered to be better for the spectral distribution or for large apertures (Boysen & Adlhart, 1987). These questions are treated in some detail by Klug & Alexander (1954). The main features, however, may also be derived by the Gaussian approximation. In practice, the function R may be obtained either by calculation from the known instrumental parameters or by measuring Bragg peaks from a perfect unstrained crystal. In the latter case, the intensity profile is given solely by the resolution function. Normalization with the Bragg intensities is also useful in order to place the diffuse-scattering intensity on an absolute scale.

In conventional single-crystal diffractometry the measured intensity is given by the convolution product of $\mathrm{d}\sigma/\mathrm{d}\Omega$ with R,

$$I(\mathbf{H}_0) = \int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\mathbf{H})[R(\mathbf{H}-\mathbf{H}_0)]\,\mathrm{d}\mathbf{H}, \tag{4.2.8.1}$$

where $\mathrm{d}\sigma/\mathrm{d}\Omega$ describes the scattering cross section for the disorder problem. In a more accurate form the mosaicity of the sample has to be included:

$$\begin{aligned} I(\mathbf{H}_0) &= \int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\mathbf{H}-\Delta\mathbf{K})[\eta(\Delta\mathbf{K})R(\mathbf{H}-\mathbf{H}_0)]\,\mathrm{d}\mathbf{H}\,\mathrm{d}(\Delta\mathbf{K}) \\ &= \int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\mathbf{H}')[R'(\mathbf{H}'-\mathbf{H}_0)]\,\mathrm{d}\mathbf{H}'. \end{aligned} \tag{4.2.8.1a}$$

$R'(\mathbf{H}'-\mathbf{H}_0) = \int \eta(\Delta\mathbf{K})R(\mathbf{H}'+\Delta\mathbf{K}-\mathbf{H}_0)\,\mathrm{d}(\Delta\mathbf{K})$. The mosaic block distribution around a most probable vector $\mathbf{K}_0$ is described by $\eta(\Delta\mathbf{K})$: $\Delta\mathbf{K} = \mathbf{K}-\mathbf{K}_0$; $\mathbf{H}' = \mathbf{H}-\Delta\mathbf{K}$.

In (4.2.8.1) all factors independent of 2θ are neglected. All intensity expressions have to be calculated from equations (4.2.8.1) or (4.2.8.1a).

The intensity variation of diffuse peaks with 2θ measured with a single detector was studied in detail by Yessik *et al.* (1973). In principle, all special cases are included there. In practice, however, some important simplifications can be made if $\mathrm{d}\sigma/\mathrm{d}\Omega$ is either very broad or very sharp compared with R, *i.e.* for Bragg peaks, sharp streaks, 'thin' diffuse layers or extended 3D diffuse peaks (Boysen & Adlhart, 1987).

In the latter case, the cross section $\mathrm{d}\sigma/\mathrm{d}\Omega$ may be treated as nearly constant over the resolution volume so that the corresponding 'Lorentz' factor is independent of 2θ:

$$L_{3\mathrm{D}} = 1. \tag{4.2.8.2}$$

For a diffuse plane within the scattering plane with very small thickness and slowly varying cross section within the plane, one derives for a point measurement in the plane

$$L_{2\mathrm{D},\parallel} = (\beta_1'^2 + \beta_2^2 + 4\eta_{2\mathrm{v}}^2\sin^2\theta)^{-1/2}, \tag{4.2.8.3}$$

exhibiting an explicit dependence on θ (β_1', β_2 and $\eta_{2\mathrm{v}}$ determine an effective vertical divergence before the sample, the divergence before the detector and the vertical mosaic spread of the sample, respectively).

In the case of *relaxed* vertical collimations $\beta_1', \beta_2 \gg \eta_{2\mathrm{v}}$

$$L_{2\mathrm{D},\parallel} = (\beta_1'^2 + \beta_2^2)^{-1/2}, \tag{4.2.8.3a}$$

i.e. again independent of θ.

Scanning across the diffuse layer in a direction perpendicular to it one obtains an integrated intensity which is also independent of 2θ. This is even true if approximations other than Gaussians are used.

If, on the other hand, an equivalent diffuse plane is positioned perpendicular to the scattering plane, the equivalent expression for $L_{2\mathrm{D},\perp}$ of a point measurement is given by

$$\begin{aligned} L_{2\mathrm{D},\perp} \simeq [&4\eta_{2\mathrm{H}}^2\sin^2\theta\cos^2\psi + \alpha_2^2\sin^2(\psi-\theta) \\ &+ \alpha_1'^2\sin^2(\psi+\theta) + 4\eta_0'\sin^2\theta\sin^2\psi \\ &- 4\alpha_1''\sin\theta\sin\psi\sin(\theta+\psi)], \end{aligned} \tag{4.2.8.4}$$

where ψ gives the angle between the vector $\mathbf{H}_0$ and the line of intersection between the diffuse plane and the scattering plane. The coefficients $\eta_{2\mathrm{H}}$, α_2, α_1', α_1'' and η are either instrumental parameters or functions of them, defining horizontal collimations and mosaic spreads. In the case of a sharp X-ray line (produced, for example, by filtering) the last two terms in equation (4.2.8.4) vanish.

The use of integrated intensities from individual scans perpendicular to the diffuse plane, now carried out within the scattering plane, again gives a Lorentz factor independent of 2θ.

In the third fundamental special case, diffuse streaking along one reciprocal direction within the scattering plane (with a narrow cross section and slowly varying intensity along the streak), the Lorentz factor for a point measurement may be expressed by the product

$$L_{1\mathrm{D},\parallel} \simeq L_{2\mathrm{D},\parallel}L_{2\mathrm{D},\perp}, \tag{4.2.8.5}$$

where ψ now defines the angle between the streak and $\mathbf{H}_0$. The integrated intensity taken from an H scan perpendicular to the streak has to be corrected by a Lorentz factor which is equal to $L_{2\mathrm{D},\parallel}$ [equation (4.2.8.3)]. In the case of a diffuse streak perpendicular to the scattering plane, a relatively complicated equation holds for the corresponding Lorentz factor (Boysen & Adlhart, 1987). Again, simpler expressions hold for integrated intensities from H scans perpendicular to the streaks. Such scans may be performed in the radial direction (corresponding to a θ–2θ scan),

$$L_{1\mathrm{D},\perp,\mathrm{rad}} = (4\eta_{2\mathrm{H}}^2 + \alpha_2^2 + \alpha_1'^2)^{-1/2}(1/\sin\theta), \tag{4.2.8.6}$$

or perpendicular to the radial direction (within the scattering plane) (corresponding to an ω scan),

$$L_{1\mathrm{D},\perp,\mathrm{per}} = (\alpha_2^2 + \alpha_1^2 + 4\eta_0' \tan^2\theta - 4\alpha_1'' \tan\theta)^{-1/2}(1/\cos\theta). \quad (4.2.8.7)$$

Note that only the radial scan yields a simple θ dependence ($\sim 1/\sin\theta$).

From these considerations it is recommended that integrated intensities from scans perpendicular to a diffuse plane or a diffuse streak should be used in order to extract the disorder cross sections. For other scan directions, which make an angle α with the intersection line (diffuse plane) or with a streak, the L factors are simply $L_{2\mathrm{D},\perp}/\sin\alpha$ and $L_{1\mathrm{D},\perp}/\sin\alpha$, respectively.

One point should be emphasized: since in a usual experiment with a single counter the integration is performed over an angle $\Delta\omega$ *via* a general $\delta\omega{:}(g\delta 2\theta)$ scan, an additional correction factor arises:

$$\Delta\omega/\Delta\mathbf{H}_\beta = \sin(\beta+\theta)/(\mathbf{k}_0 \sin 2\theta). \quad (4.2.8.8)$$

β is the angle between $\mathbf{H}_0$ and the scan direction $\mathbf{H}_\beta$, and $g = (\tan\beta + \tan\theta)/(2\tan\theta)$ defines the coupling ratio between the rotation of the crystal around a vertical axis and the rotation of the detector shaft. The so-called 1:2 and ω-scan techniques are most frequently used, where $\beta = 0$ and 90°, respectively.

White-beam techniques: Techniques for the measurement of diffuse scattering using a *white* spectrum are common in neutron diffraction. Owing to the relatively low velocity of thermal or cold neutrons, TOF methods in combination with time-resolving detector systems placed at a fixed angle 2θ allow for a simultaneous recording along a radial direction through the origin of reciprocal space (see, *e.g.*, Turberfield, 1970; Bauer *et al.*, 1975). The scan range is limited by the Ewald spheres corresponding to $\lambda_{\max}$ and $\lambda_{\min}$, respectively. With several such detector systems placed at different angles or a 2D detector several scans may be carried out simultaneously during one neutron pulse.

An analogue of neutron TOF diffractometry in the X-ray case is a combination of a white source of X-rays and an energy-dispersive detector. This technique, which has been known in principle for a long time, suffered from relatively weak white sources. With the development of high-power X-ray generators and synchrotron sources this method has now become highly interesting. Its use in diffuse-scattering work (in particular, the effects on resolution) is discussed by Harada *et al.* (1984).

Some examples of neutron instruments dedicated to diffuse scattering are the diffractometers D7 at the Institut Laue–Langevin (ILL, Grenoble), DNS at the Forschungsneutronenquelle Heinz Maier-Leibnitz (FRM-II, Munich), G4-4 at the Laboratoire Léon Brillouin (LLB, Saclay) or SXD at ISIS, Rutherford Appleton Laboratory (RAL, UK). The instrument DNS uses a polarization analysis for diffuse scattering (Schweika & Böni, 2001; Schweika, 2003) and the SXD diffractometer is a TOF instrument using a (pulsed) white beam (Keen & Nield, 2004). All these instruments are equipped with banks of detectors. The single-crystal diffractometer D19 at the ILL, equipped with a multiwire area detector, is also suitable for collecting diffuse data. The flat-cone machine E2 at the Hahn–Meitner Institut (HMI, Berlin) is equipped with a bank of area detectors, has an option to record higher-order layers and can also be operated in an 'elastic' mode with a multicrystal analyser. The instrument D10 at the ILL is a versatile instrument which can be operated as a low-background two-axis or three-axis diffractometer with several further options. Neutron diffractometers that have recently become operational are BIX-3 at the Japan Atomic Energy Research Institute (JAERI, Japan), LADI at the ILL, where neutron-sensitive IPs are used for macromolecular work (for a comparison see Niimura *et al.*, 2003), and RESI at the Forschungsneutronenquelle Heinz Maier-Leibnitz (FRM-II, Munich) for common solid-state investigations. Information about all these instruments can be found in the respective handbooks or on the websites of the facilities.

4.2.8.2. Powders and polycrystals

The diffuse background in powder diagrams also contains valuable information about disorder. Only in very simple cases can a model be deduced from a powder pattern alone, but a refinement of a known disorder model can favourably be carried out, *e.g.* the temperature dependence may be studied. On account of the intensity integration, the ratio of diffuse intensity to Bragg intensity is enhanced in a powder pattern. Moreover, a powder pattern contains, in principle, all the information about the sample and might thus reveal more than single-crystal work. However, in powder-diffractometer experiments preferred orientations and textures could lead to a complete misidentification of the problem. Single-crystal experiments are generally preferable in this respect. Nevertheless, high-resolution powder investigations may give quick supporting information, *e.g.* about superlattice peaks, split reflections, lattice strains, domain-size effects, lattice-constant changes related to a disorder effect *etc*.

Evaluation of diffuse-scattering data from powder diffraction follows the same theoretical formulae developed for the determination of the radial distribution function for glasses and liquids. The final formula for random distributions may be given as (Fender, 1973)

$$I_D^p = \{\langle|F(\mathbf{H})|^2\rangle - |\langle F(\mathbf{H})\rangle|^2\} \sum_i s_i \sin(2\pi H r_i)/(2\pi H r_i). \quad (4.2.8.9)$$

s_i represents the number of atoms at distance r_i from the origin. An equivalent expression for a substitutional binary alloy is

$$I_D^p = \alpha(1-\alpha)\{|f_2(\mathbf{H}) - f_2(\mathbf{H})|^2\} \sum_i s_i \sin(2\pi H r_i)/(2\pi H r_i). \quad (4.2.8.10)$$

A quantitative calculation of a diffuse background is also helpful in combination with Rietveld's method (1969) for refining an averaged structure by fitting Bragg data. In particular, for highly anisotropic diffuse phenomena characteristic asymmetric line shapes occur.

The calculation of these line shapes is treated in the literature, mostly neglecting the instrumental resolution (see, *e.g.*, Warren, 1941; Wilson, 1949; Jones, 1949; and de Courville-Brenasin *et al.*, 1981). This is not justified if the variation of the diffuse intensity becomes comparable with that of the resolution function, as is often the case in neutron diffraction. The instrumental resolution may be incorporated using the resolution function of a powder instrument (Caglioti *et al.*, 1958). A detailed analysis of diffuse peaks is given by Yessik *et al.* (1973) and the equivalent considerations for diffuse planes and streaks are discussed by Boysen (1985). The case of 3D random disorder (incoherent neutron scattering, monotonous Laue scattering, averaged TDS, multiple scattering or short-range-order modulations) is treated by Sabine & Clarke (1977).

In polycrystalline samples the cross section has to be averaged over all orientations:

$$\frac{\mathrm{d}\sigma_\mathrm{p}}{\mathrm{d}\Omega}(\mathbf{H}') = \frac{n_\mathrm{c}}{H^2}\int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\mathbf{H}') R'(|\mathbf{H}'| - |\mathbf{H}_0|)\,\mathrm{d}\mathbf{H}', \quad (4.2.8.11)$$

where n_c is number of crystallites in the sample; this averaged cross section enters the relevant expressions for the convolution product with the resolution function.

A general intensity expression may be written as (Boysen, 1985)

$$I_n(H_0) = P \sum_T m(T) A_n \Phi_n(H_0, T). \quad (4.2.8.12)$$

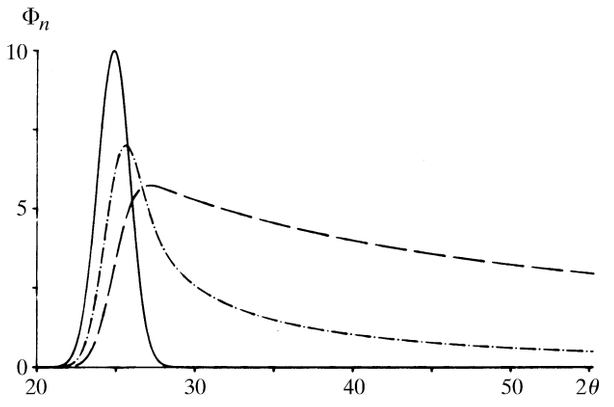


Fig. 4.2.8.2. Line profiles in powder diffraction for diffuse peaks (full line), continuous streaks (dot–dash lines) and continuous planes (broken lines). For explanation see text.

P denotes a scaling factor that depends on the instrumental luminosity, T is the shortest distance to the origin of the reciprocal lattice, $m(T)$ is the corresponding symmetry-induced multiplicity, A_n contains the structure factor of the structural units and the type of disorder, and Φ_n describes the characteristic modulation of the diffuse phenomenon of dimension n in the powder pattern. These expressions are given below with the assumption of Gaussian line shapes of width D for the narrow extension(s). The formulae depend on a factor $M = A_{1/2}(4k_1^2 - H_0^2)/(32\ln 2)$, where $A_{1/2}$ describes the dependence of the Bragg peaks on the instrumental parameters U, V and W (see Caglioti *et al.*, 1958),

$$A_{1/2}^2 = U\tan^2\theta + V\tan\theta + W, \qquad (4.2.8.13)$$

and $k_1 = 1/\lambda$.

(*a*) *Isotropic diffuse peak around T*

$$\Phi_0 = [2\pi(M^2 + D^2)]^{-1/2}(1/T^2) \times \exp\{-(H_0 - T)^2/2(M^2 + D^2)\}. \qquad (4.2.8.14)$$

The moduli $|\mathbf{H}_0|$ and T enter the exponential, *i.e.* the variation of $d\sigma/d\Omega$ along $|\mathbf{H}_0|$ is essential. For broad diffuse peaks ($M \ll D$) the angular dependence is due to $1/T^2$, *i.e.* proportional to $1/\sin^2\theta$. This result is valid for diffuse peaks of any shape.

(*b*) *Diffuse streak*

$$\Phi_1 = [2\pi(M^2 + D^2)]^{-1/2} \int_{-\infty}^{\infty} (T^2 + H^2)^{-1/2} \times \exp\{-H_0 - (T^2 + H^2)^{1/2}/2(M^2 + D^2)\}\,dH. \qquad (4.2.8.15)$$

The integral has to be evaluated numerically. If $(M^2 + D^2)$ is not too large, the term $1/(T^2 + H^2)$ varies only slowly compared to the exponential term and may be kept outside the integral, setting it approximately to $1/H_0^2$.

(*c*) *Diffuse plane (with $R^2 = H_x^2 + H_y^2$)*

$$\Phi_2 = (M^2 + D^2)^{-1/2} \int R^2/(T^2 + R^2) \times \exp\{-H_0 - (T^2 + R^2)^{1/2}/2(M^2 + D^2)\}\,dR. \qquad (4.2.8.16)$$

With the same approximation as in (*b*) the expression may be simplified to

$$\Phi_2 = \pi/H_0[1 - \mathrm{erf}\{(T - H_0)/[2(M^2 + D^2)]^{1/2}\} + 1/H_0^2[2\pi(M^2 + D^2)]^{1/2} \times \exp\{-(H_0 - T)^2/2(M^2 + D^2)\}]. \qquad (4.2.8.17)$$

(*d*) *Slowly varying diffuse scattering in three dimensions*

Φ_3 = constant. Consequently, the intensity is directly proportional to the cross section. The characteristic functions Φ_0, Φ_1 and Φ_2 are shown in Fig. 4.2.8.2 for equal values of T and D. Note the relative peak shifts and the high-angle tail.

4.2.8.3. Total diffraction pattern

As mentioned in Section 4.2.7.3.2, the atomic pair-distribution function (PDF), which is classically used for the analysis of the atomic distributions in liquids, melts or amorphous samples, can also be used to gain an understanding of disorder in crystals. The PDF is the Fourier transform of the total scattering. The measurement of total scattering is basically similar to recording X-ray or neutron powder patterns. The success of the method depends, however, decisively on various factors: (i) The availability of a large data set, *i.e.* reliable intensities up to high H values, in order to get rid of truncation ripples, which heavily influence the interpretation. Currently, values of $H_{\max}$ of more than 7 Å^{-1} can be achieved either with synchrotron X-rays [at the European Synchrotron Radiation Facility (ESRF) or Cornell High Energy Synchrotron Source (CHESS)] or with neutrons from reactors (*e.g.* instrument D4 at the ILL) and spallation sources [at ISIS or at LANSCE (instrument NPDF)]. (ii) High H resolution. (iii) High intensities, in particular at high H values. (iv) Low background of any kind which does not originate from the sample. High-quality intensities are therefore to be extracted from the raw data by taking care of an adequate absorption correction, correction of multiple-scattering effects, separation of inelastically scattered radiation (*e.g.* Compton scattering) and careful subtraction of 'diffuse' background which is not the 'true' diffuse scattering from the sample. These conditions are in practice rather demanding. A further detailed discussion is beyond the scope of this chapter, but a more thorough discussion is given, *e.g.*, by Egami (2004), and some examples are given by Egami & Billinge (2003).

References

Adlhart, W. (1981). *Diffraction by crystals with planar domains*. *Acta Cryst.* A**37**, 794–801.

Amorós, J. L. & Amorós, M. (1968). *Molecular Crystals: Their Transforms and Diffuse Scattering*. New York: John Wiley.

Arndt, U. W. (1986*a*). *X-ray position sensitive detectors*. *J. Appl. Cryst.* **19**, 145–163.

Arndt, U. W. (1986*b*). *The collection of single crystal diffraction data with area detectors*. *J. Phys. (Paris) Colloq.* **47**(C5), 1–6.

Aroyo, M. I., Boysen, H. & Perez-Mato, J. M. (2002). *Inelastic neutron scattering selection rules for phonons: application to leucite phase transitions*. *Appl. Phys. A*, **74**, S1043–S1048.

Axe, J. D. (1980). *Debye–Waller factors for incommensurate structures*. *Phys. Rev. B*, **21**, 4181–4190.

Bardhan, P. & Cohen, J. B. (1976). *X-ray diffraction study of short-range-order structure in a disordered Au_3Cu alloy*. *Acta Cryst.* A**32**, 597–614.

Bauer, G., Seitz, E. & Just, W. (1975). *Elastic diffuse scattering of neutrons as a tool for investigation of non-magnetic point defects*. *J. Appl. Cryst.* **8**, 162–175.

Bauer, G. S. (1979). *Diffuse elastic neutron scattering from nonmagnetic materials*. In *Treatise on Materials Science and Technology*, Vol. 15, edited by G. Kostorz, pp. 291–336. New York: Academic Press.

Berthold, T. & Jagodzinski, H. (1990). *Analysis of the distribution of impurities in crystals by anomalous X-ray scattering*. *Z. Kristallogr.* **193**, 85–100.

Bessière, M., Lefebvre, S. & Calvayrac, Y. (1983). *X-ray diffraction study of short-range order in a disordered Au_3Cu alloy*. *Acta Cryst.* B**39**, 145–153.

Beyeler, H. U., Pietronero, L. & Strässler, S. (1980). *Configurational model for a one-dimensional ionic conductor*. *Phys. Rev. B*, **22**, 2988–3000.

Billinge, S. J. L. & Egami, T. (1993). *Short-range atomic structure of $Nd_{2-x}Ce_xCuO_{4-y}$ determined by real-space refinement of neutron-powder-diffraction data. Phys. Rev B*, **47**, 14386–14406.

Billinge, S. J. L. & Thorpe, M. F. (1998). *Local Structure from Diffraction*. New York: Plenum Press.

Boissieu, M. de, Boudard, M., Hennion, M., Bellissent, R., Kycia, S., Goldman, A., Janot, C. & Audier, M. (1995). *Diffuse scattering and phason elasticity in the AlPdMn icosahedral phase. Phys. Rev. Lett.* **75**, 89–92.

Borie, B. & Sparks, C. J. (1971). *The interpretation of intensity distributions from disordered binary alloys. Acta Cryst.* A**27**, 198–201.

Boysen, H. (1985). *Analysis of diffuse scattering in neutron powder diagrams. Applications to glassy carbon. J. Appl. Cryst.* **18**, 320–325.

Boysen, H. (1995). *Diffuse scattering by domains and domain walls*. *Phase Transit.* **55**, 1–16.

Boysen, H. (1997). *Suppression of subsidiary maxima in computer simulations of diffraction intensities*. *Z. Kristallogr*. **212**, 634–636.

Boysen, H. (2007). *Coherence effects in the scattering from domain structures*. *J. Phys. Condens. Matter*, **19**, 270526–270537.

Boysen, H. & Adlhart, W. (1987). *Resolution corrections in diffuse scattering experiments. J. Appl. Cryst.* **20**, 200–209.

Boysen, H. & Frey, F. (1998). *Diffuse scattering: X-rays vs. neutrons. Phase Transit.* **67**, 197–217.

Boysen, H., Frey, F., Schrader, H. & Eckold, G. (1991). *On the proto-to-clino-/orthoenstatite phase transformation: single-crystal X-ray and inelastic neutron investigation*. *Phys. Chem. Mineral.* **17**, 629–635.

Bradley, C. J. & Cracknell, A. P. (1972). *The Mathematical Theory of Symmetry in Solids*, pp. 51–76. Oxford: Clarendon Press.

Brämer, R. (1975). *Statistische Probleme der Theorie des Parakristalls. Acta Cryst.* A**31**, 551–560.

Brämer, R. & Ruland, W. (1976). *The limitations of the paracrystalline model of disorder. Macromol. Chem.* **177**, 3601–3617.

Bubeck, E. & Gerold, V. (1984). *An X-ray investigation in the small and wide angle range from G.P.I. zones in Al–Cu*. In *Microstructural Characterization of Materials by Non-Microscopical Techniques*, edited by N. H. Andersen, M. Eldrup, N. Hansen, R. J. Jensen, T. Leffers, H. Lilholt, O. B. Pedersen & B. N. Singh. Roskilde: Risø National Laboratory.

Burandt, B., Komorek, M., Schnabel, W., Press, W. & Boysen, H. (1992). *High resolution X-ray investigations on the supersatellite reflections of labradorite. Z. Kristallogr.* **200**, 141–156.

Butler, B. D., Haeffner, D. R., Lee, P. L. & Welberry, T. R. (2000). *High-energy X-ray diffuse scattering using Weissenberg flat-cone geometry*. *J. Appl. Cryst.* **33**, 1046–1050.

Butler, B. D. & Welberry, T. R. (1992). *Calculation of diffuse scattering from simulated disordered crystals: a comparison with optical transforms*. *J. Appl. Cryst.* **25**, 391–399.

Caglioti, G., Paoletti, A. & Ricci, R. (1958). *Choice of collimators for a crystal spectrometer for neutron diffraction. Nucl. Instrum. Methods*, **3**, 223–226.

Campbell, B. J., Welberry, T. R., Broach, R. W., Hong, H. & Cheetham, A. K. (2004). *Elucidation of zeolite microstructure by synchrotron X-ray diffuse scattering. J. Appl. Cryst.* **37**, 187–192.

Cenedese, P., Bley, F. & Lefebvre, S. (1984). *Diffuse scattering in disordered ternary alloys: neutron measurements of local order in a stainless steel $Fe_{0.56}Cr_{0.21}Ni_{0.23}$. Acta Cryst.* A**40**, 228–240.

Collongues, R., Fayard, M. & Gautier, F. (1977). *Ordre et désordre dans les solides. J. Phys. (Paris) Colloq.* **38**(C7), Suppl.

Comes, R. & Shirane, G. (1979). *X-ray and neutron scattering from one-dimensional conductors.* In *Highly Conducting One Dimensional Solids*, edited by J. T. Devreese, R. P. Evrard & V. E. Van Doren, ch. 2. New York: Plenum.

Conradi, E. & Müller, U. (1986). *Fehlordnung bei Verbindungen mit Schichtstrukturen. II. Analyse der Fehlordnung in Wismuttriiodid. Z. Kristallogr.* **176**, 263–269.

Convert, P. & Forsyth, J. B. (1983). *Position-Sensitive Detectors of Thermal Neutrons.* London: Academic Press.

Cooper, M. J. & Nathans, R. (1968*a*). *The resolution function in neutron diffractometry. II. The resolution function of a conventional two-crystal neutron diffractometer for elastic scattering. Acta Cryst.* A**24**, 481–484.

Cooper, M. J. & Nathans, R. (1968*b*). *The resolution function in neutron diffractometry. III. Experimental determination and properties of the elastic two-crystal resolution function. Acta Cryst.* A**24**, 619–624.

Courville-Brenasin, J. de, Joyez, G. & Tchoubar, D. (1981). *Méthode d'ajustement automatique entre courbes experiméntale et calculée dans les diagrammes de diffraction de poudre. Cas des solides à structure lamellaire. I. Développement de la méthode. J. Appl. Cryst.* **14**, 17–23.

Cowley, J. M. (1950*a*). *X-ray measurement of order in single crystals of Cu_3Au. J. Appl. Phys.* **21**, 24–36.

Cowley, J. M. (1950*b*). *An approximate theory of order in alloys. Phys. Rev.* **77**, 664–675.

Cowley, J. M. (1976*a*). *Diffraction by crystals with planar faults. I. General theory. Acta Cryst.* A**32**, 83–87.

Cowley, J. M. (1976*b*). *Diffraction by crystals with planar faults. II. Magnesium fluorogermanate. Acta Cryst.* A**32**, 88–91.

Cowley, J. M. (1981). *Diffraction Physics*, 2nd ed. Amsterdam: North-Holland.

Cowley, J. M. & Au, A. Y. (1978). *Diffraction by crystals with planar faults. III. Structure analysis using microtwins. Acta Cryst.* A**34**, 738–743.

Cowley, J. M., Cohen, J. B., Salamon, M. B. & Wuensch, B. J. (1979). *Modulated Structures. AIP Conference Proceedings*, No. 53. New York: AIP.

Dederichs, P. H. (1973). *The theory of diffuse X-ray scattering and its application to the study of point defects and their clusters. J. Phys. F*, **3**, 471–496.

Dolling, G., Powell, B. M. & Sears, V. F. (1979). *Neutron diffraction study of the plastic phases of polycrystalline SF_6 and CBr_4. Mol. Phys.* **37**, 1859–1883.

Dorner, B. & Comes, R. (1977). *Phonons and structural phase transformation.* In *Dynamics of Solids and Liquids by Neutron Scattering*, edited by S. Lovesey & T. Springer, ch. 3. *Topics in Current Physics*, Vol. 3. Berlin: Springer.

Dorner, C. & Jagodzinski, H. (1972). *Entmischung im System SnO_2–TiO_2. Krist. Tech.* **7**, 427–444.

Dubernat, J. & Pezerat, H. (1974). *Fautes d'empilement dans les oxalates dihydratés des métaux divalents de la série magnésienne (Mg, Fe, Co, Ni, Zn, Mn). J. Appl. Cryst.* **7**, 387–393.

Edwards, O. S. & Lipson, H. (1942). *Imperfections in the structure of cobalt. I. Experimental work and proposed structure. Proc. R. Soc. London Ser. A*, **180**, 268–277.

Egami, T. (1990). *Atomic correlations in non-periodic matter*. *Mater. Trans. JIM*, **31**, 163–176.

Egami, T. (2004). *Local crystallography of crystals with disorder. Z. Kristallogr.* **219**, 122–129.

Egami, T. & Billinge, S. J. L. (2003). *Underneath the Bragg Peaks: Structural Analysis of Complex Materials.* Oxford: Pergamon Press.

Emery, V. J. & Axe, J. D. (1978). *One-dimensional fluctuations and the chain-ordering transformation in $Hg_{3-\delta}AsF_6$. Phys. Rev. Lett.* **40**, 1507–1511.

Endres, H., Pouget, J. P. & Comes, R. (1982). *Diffuse X-ray scattering and order–disorder effects in the iodine chain compounds N,N′-diethyl-N,N′-dihydrophenazinium iodide, $E_2PI_{1.6}$ and N,N′-dibenzyl-N,N′-dihydrophenazinium iodide, $B_2PI_{1.6}$. J. Phys. Chem. Solids*, **43**, 739–748.

Epstein, J., Welberry, T. R. & Jones, R. (1982). *Analysis of the diffuse X-ray scattering from substitutionally disordered molecular crystals of monoclinic 9-bromo-10-methylanthracene. Acta Cryst.* A**38**, 611–618.

Estermann, M. & Steurer, W. (1998). *Diffuse scattering data acquisition techniques. Phase Transit.* **67**, 165–195.

Fender, B. E. F. (1973). *Diffuse scattering and the study of defect solids.* In *Chemical Applications of Thermal Neutron Scattering*, edited by B. T. M. Willis, ch. 11. Oxford University Press.

Flack, H. D. (1970). *Short-range order in crystals of anthrone and in mixed crystals of anthrone–anthraquinone. Philos. Trans. R. Soc. London Ser. A*, **266**, 583–591.

Fontaine, D. de (1972). *An analysis of clustering and ordering in multicomponent solid solutions. I. Stability criteria. J. Phys. Chem. Solids*, **33**, 297–310.

Fontaine, D. de (1973). *An analysis of clustering and ordering in multicomponent solid solutions. II. Fluctuations and kinetics. J. Phys. Chem. Solids*, **34**, 1285–1304.

Forst, R., Jagodzinski, H., Boysen, H. & Frey, F. (1987). *Diffuse scattering and disorder in urea inclusion compounds $OC(NH_2)_2 + C_nH_{2n+2}$. Acta Cryst.* B**43**, 187–197.

Fouret, P. (1979). *Diffuse X-ray scattering by orientationally disordered solids.* In *The Plastically Crystalline State*, edited by J. N. Sherwood, ch. 3. New York: John Wiley.

Frey, F. (1997). *Diffuse scattering by domain structures.* In *Local Structure from Diffraction*, edited by S. J. L. Billinge & M. F. Thorpe. New York: Plenum Press.

Frey, F. (2003). *Disorder diffuse scattering of crystals and quasicrystals.* In *Particle Scattering, X-ray Diffraction, and Microstructure of Solids and Liquids*, edited by M. Ristig & K. Gernoth. Heidelberg: Springer.

Frey, F. & Boysen, H. (1981). *Disorder in cobalt single crystals. Acta Cryst.* A**37**, 819–826.

Frey, F., Jagodzinski, H. & Steger, W. (1986). *On the phase transformation zinc blende to wurtzite. Bull. Minéral. Crystallogr.* **109**, 117–129.

Frey, F. & Weidner, E. (2002). *Diffraction by domains in decagonal Al-Co-Ni quasicrystals. J. Alloys Compd.* **342**, 105–109.

Frey, F. & Weidner, E. (2003). *Disorder in decagonal quasicrystals. Z. Kristallogr.* **218**, 160–169.

Fultz, B. & Howe, J. M. (2002). *Transmission Electron Microscopy and Diffractometry of Materials*, ch. 9. Heidelberg: Springer.

Gehlen, P. & Cohen, J. B. (1965). *Computer simulation of the structure associated with local order in alloys. Phys. Rev. A*, **139**, 844–855.

Georgopoulos, P. & Cohen, J. B. (1977). *The determination of short range order and local atomic displacements in disordered binary solid solutions. J. Phys. (Paris) Colloq.* **38**(C7), 191–196.

Gerlach, P., Schärpf, O., Prandl, W. & Dorner, B. (1982). *Separation of the coherent and incoherent scattering of C_2Cl_6 by polarization analysis. J. Phys. (Paris) Colloq.* **43**(C7), 151–157.

Gerold, V. (1954). *Röntgenographische Untersuchungen über die Aushärtung einer Aluminium–Kupfer-Legierung mit Kleinwinkel-Schwenkaufnahmen. Z. Metallkd.* **45**, 593–607.

Gibbons, P. C. & Kelton, K. F. (1993). *Arcs of diffuse scattering in icosahedral glass models. J. Non-Cryst. Solids*, **153&154**, 165–171.

Goff, J. P., Hayes, W., Hull, S., Hutching, M. T. & Clausen, K. N. (1999). *Defect structure of yttria-stabilized zirconia and its influence on the ionic conductivity at elevated temperatures. Phys. Rev. B*, **59**, 14202–14219.

Goldman, A. I., Guryan, C. A., Stephens, P. W., Parsey, J. M. Jr, Aeppli, G., Chen, H. S. & Gayle, F. W. (1988). *Diffuse scattering from the icosahedral phase alloys. Phys. Rev. Lett.* **61**, 1962–1965.

Gragg, J. E., Hayakawa, M. & Cohen, J. B. (1973). *Errors in qualitative analysis of diffuse scattering from alloys. J. Appl. Cryst.* **6**, 59–66.

Guinier, A. (1942). *Le mécanisme de la précipitation dans un cristal de solution solide métallique. Cas des systémes aluminum–cuivre et aluminum–argent. J. Phys. Radium*, **8**, 124–136.

Guinier, A. (1963). *X-ray Diffraction in Crystals, Imperfect Solids and Amorphous Bodies.* San Francisco: Freeman.

Halla, F., Jagodzinski, H. & Ruston, W. R. (1953). *One-dimensional disorder in dodecahydrotriphenylene, $C_{18}H_{24}$. Acta Cryst.* **6**, 478–488.

Harada, J., Iwata, H. & Ohshima, K. (1984). *A new method for the measurement of X-ray diffuse scattering with a combination of an energy dispersive detector and a source of white radiation. J. Appl. Cryst.* **17**, 1–6.

Harburn, G., Miller, J. S. & Welberry, T. R. (1974). *Optical-diffraction screens containing large numbers of apertures. J. Appl. Cryst.* **7**, 36–38.

Harburn, G., Taylor, C. A. & Welberry, T. R. (1975). *An Atlas of Optical Transforms.* London: Bell.

Hashimoto, S. (1974). *Correlative microdomain model for short range ordered alloy structures. I. Diffraction theory. Acta Cryst.* A**30**, 792–798.

Hashimoto, S. (1981). *Correlative microdomain model for short range ordered alloy structures. II. Application to special cases. Acta Cryst.* A**37**, 511–516.

Hashimoto, S. (1983). *Correlative microdomain model for short range ordered alloy structures. III. Analysis for diffuse scattering from quenched CuAu alloy. Acta Cryst.* A**39**, 524–530.

Hashimoto, S. (1987). *Intensity expression for short-range order diffuse scattering with ordering energies in a ternary alloy system. J. Appl. Cryst.* **20**, 182–186.

Haubold, H. G. (1975). *Measurement of diffuse X-ray scattering between reciprocal lattice points as a new experimental method in determining interstitial structures. J. Appl. Cryst.* **8**, 175–183.

Hayakawa, M. & Cohen, J. B. (1975). *Experimental considerations in measurements of diffuse scattering. Acta Cryst.* A**31**, 635–645.

Hendricks, S. B. & Teller, E. (1942). *X-ray interference in partially ordered layer lattices. J. Chem. Phys.* **10**, 147–167.

Hohlwein, D., Hoser, A. & Prandl, W. (1986). *Orientational disorder in cubic $CsNO_2$ by neutron powder diffraction. Z. Kristallogr.* **177**, 93–102.

Hosemann, R. (1975). *Micro paracrystallites and paracrystalline superstructures. Macromol. Chem.* Suppl. 1, pp. 559–577.

Hosemann, R. & Bagchi, S. N. (1962). *Direct Analysis of Diffraction by Matter.* Amsterdam: North-Holland.

Iizumi, M. (1973). *Lorentz factor in single crystal neutron diffraction. Jpn. J. Appl. Phys.* **12**, 167–172.

Ishii, T. (1983). *Static structure factor of Frenkel–Kontorova-systems at high temperatures. Application to K-hollandite. J. Phys. Soc. Jpn*, **52**, 4066–4073.

Jagodzinski, H. (1949*a*). *Eindimensionale Fehlordnung und ihr Einfluss auf die Röntgeninterferenzen. I. Berechnung des Fehlordnungsgrades aus den Röntgeninterferenzen. Acta Cryst.* **2**, 201–208.

Jagodzinski, H. (1949*b*). *Eindimensionale Fehlordnung und ihr Einfluss auf die Röntgeninterferenzen. II. Berechnung der fehlgeordneten dichtesten Kugelpackungen mit Wechsel wirkungen der Reichweite 3. Acta Cryst.* **2**, 208–214.

Jagodzinski, H. (1949*c*). *Eindimensionale Fehlordnung und ihr Einfluss auf die Röntgeninterferenzen. III. Vergleich der Berechungen mit experimentellen Ergebnissen. Acta Cryst.* **2**, 298–304.

Jagodzinski, H. (1954). *Der Symmetrieeinfluss auf den allgemeinen Lösungsansatz eindimensionaler Fehlordnungsprobleme. Acta Cryst.* **7**, 17–25.

Jagodzinski, H. (1963). *On disorder phenomena in crystals.* In *Crystallography and Crystal Perfection*, edited by G. N. Ramachandran, pp. 177–188. London: Academic Press.

Jagodzinski, H. (1964*a*). *Allgemeine Gesichtspunkte für die Deutung diffuser Interferenzen von fehlgeordneten Kristallen.* In *Advances in Structure Research by Diffraction Methods*, Vol. 1, edited by R. Brill & R. Mason, pp. 167–198. Braunschweig: Vieweg.

Jagodzinski, H. (1964*b*). *Diffuse disorder scattering by crystals.* In *Advanced Methods of Crystallography*, edited by G. N. Ramachandran, pp. 181–219. London: Academic Press.

Jagodzinski, H. (1968). *Fokussierende Monochromatoren für Einkristallverfahren? Acta Cryst.* B**24**, 19–23.

Jagodzinski, H. (1972). *Transformation from cubic to hexagonal silicon carbide as a solid state reaction. Kristallografiya*, **16**, 1235–1246.

Jagodzinski, H. (1987). *Diffuse X-ray scattering from crystals.* In *Progress in Crystal Growth and Characterization*, edited by P. Krishna, pp. 47–102. Oxford: Pergamon Press.

Jagodzinski, H. & Haefner, K. (1967). *On order–disorder in ionic non-stoichiometric crystals. Z. Kristallogr.* **125**, 188–200.

Jagodzinski, H. & Hellner, E. (1956). *Die eindimensionale Phasenumwandlung des $RhSn_2$. Z. Kristallogr.* **107**, 124–149.

Jagodzinski, H. & Korekawa, M. (1965). *Supersatelliten im Beugungsbild des Labradorits $(Ca_2Na)(Si_2Al)_2O_8$. Naturwissenschaften*, **52**, 640–641.

Jagodzinski, H. & Korekawa, M. (1973). *Diffuse X-ray scattering by lunar minerals. Geochim. Cosmochim. Acta Suppl.* 4, **1**, 933–951.

Jagodzinski, H. & Laves, R. (1947). *Über die Deutung der Entmischungsvorgänge in Mischkristallen unter besonderer Berücksichtigung der Systeme Aluminium–Kupfer und Aluminium–Silber. Z. Metallkd.* **40**, 296–305.

Jagodzinski, H. & Penzkofer, B. (1981). *A new type of satellite in plagioclases. Acta Cryst.* A**37**, 754–762.

James, R. W. (1954). *The Optical Principles of Diffraction of X-rays*, ch. X. London: Bell.

Janner, A. & Janssen, T. (1980*a*). *Symmetry of incommensurate crystal phases. I. Commensurate basic structures. Acta Cryst.* A**36**, 399–408.

Janner, A. & Janssen, T. (1980*b*). *Symmetry of incommensurate crystal phases. II. Incommensurate basic structures. Acta Cryst.* A**36**, 408–415.

Janssen, T. & Janner, A. (1987). *Incommensurability in crystals. Adv. Phys.* **36**, 519–624.

Jaric, M. V. & Nelson, D. R. (1988). *Diffuse scattering from quasicrystals. Phys. Rev. B*, **37**, 4458–4472.

Jefferey, J. W. (1953). *Unusual X-ray diffraction effects from a crystal of wollastonite. Acta Cryst.* **6**, 821–826.

Jones, R. C. (1949). *X-ray diffraction by randomly oriented line gratings. Acta Cryst.* **2**, 252–257.

Kaiser-Bischoff, I., Boysen, H., Frey, F., Hoffmann, J.-U., Hohlwein, D. & Lerch, M. (2005). *The defect structure of Y- and N-doped zirconia. J. Appl. Cryst.* **38**, 139–146.

Kakinoki, J. & Komura, Y. (1954). *Intensity of X-ray diffraction by a one-dimensionally disordered crystal. I. General derivation in the cases of 'Reichweite' s = 0 and s = 1. J. Phys. Soc. Jpn*, **9**, 169–183.

Kakinoki, J. & Komura, Y. (1965). *Diffraction by a one-dimensionally disordered crystal. I. The intensity equation. Acta Cryst.* **19**, 137–147.

Kalning, M., Dorna, V., Press, W., Kek, S. & Boysen, H. (1997). *Profile analysis of the supersatellite reflections in labradorite – A synchrotron X-ray diffraction study. Z. Kristallogr.* **212**, 545–549.

Keen, D. A. & Nield, V. M. (2004). *Diffuse Neutron Scattering from Crystalline Materials.* Oxford University Press.

Kitaigorodsky, A. I. (1984). *Mixed Crystals. Springer Series in Solid State Science*, Vol. 33, chs. 6.4, 6.5, 8.4. Berlin: Springer.

Klug, H. P. & Alexander, L. E. (1954). *X-ray Diffraction Procedures.* New York: John Wiley.

Korekawa, M. (1967). *Theorie der Satellitenreflexe.* Habilitationsschrift der Naturwissenschaftlichen Fakultät der Universität München, Germany.

Korekawa, M. & Jagodzinski, H. (1967). *Die Satellitenreflexe des Labradorits. Schweiz. Mineral. Petrogr. Mitt.* **47**, 269–278.

Krivoglaz, M. A. (1969). *Theory of X-ray and Thermal Neutron Scattering by Real Crystals.* Part I. New York: Plenum.

Krivoglaz, M. A. (1996*a*). *Diffuse Scattering of X-rays and Neutrons by Fluctuations*. Berlin: Springer.

Krivoglaz, M. A. (1996*b*). *X-ray and Neutron Diffraction in Nonideal Crystals*. Berlin: Springer.

Lechner, R. E. & Riekel, C. (1983). *Application of neutron scattering in chemistry.* In *Neutron Scattering and Muon Spin Rotation. Springer Tracts in Modern Physics*, Vol. 101, edited by G. Höhler, pp. 1–84. Berlin: Springer.

Lefebvre, J., Fouret, R. & Zeyen, C. (1984). *Structure determination of sodium nitrate near the order–disorder phase transition. J. Phys. (Paris)*, **45**, 1317–1327.

Lei, J., Wang, R., Hu, Ch. & Ding, D. (1999). *Diffuse scattering from decagonal quasicrystals. Phys. Rev. B*, **59**, 822–828.

McGreevy, R. L. (2001). *Reverse Monte Carlo modelling. J. Phys. Condens. Matter*, **13**, R877–R913.

McGreevy, R. L. & Pusztai, L. (1988). *Reverse Monte Carlo simulation: a new technique for the determination of disordered structures. Mol. Simul.* **1**, 359–367.

Mardix, S. & Steinberger, I. T. (1970). *Tilt and structure transformation in ZnS. J. Appl. Phys.* **41**, 5339–5341.

Martorana, A., Marigo, A., Toniolo, L. & Zenetti, R. (1986). *Stacking faults in the β-form of magnesium dichloride. Z. Kristallogr.* **176**, 1–12.

Matsubara, E. & Georgopoulos, P. (1985). *Diffuse scattering measurements with synchrotron radiation: instrumentation and techniques. J. Appl. Cryst.* **18**, 377–383.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *Equation of state calculations by fast computing machines. J. Chem. Phys.* **21**, 1087–1092.

Micu, A. M. & Smith, J. C. (1995). *SERENA: a program for calculating X-ray diffuse scattering intensities from molecular dynamics trajectories. Comput. Phys. Commun.* **91**, 331–338.

More, M., Lefebvre, J. & Hennion, B. (1984). *Quasi-elastic coherent neutron scattering in the disordered phase of CBr_4. Experimental evidence of local order and rotational dynamics of molecules. J. Phys. (Paris)*, **45**, 303–307.

More, M., Lefebvre, J., Hennion, B., Powell, B. M. & Zeyen, C. (1980). *Neutron diffuse scattering in the disordered phase of CBr_4. I. Experimental. Elastic and quasi-elastic coherent scattering in single crystals. J. Phys. C*, **13**, 2833–2846.

Moss, D. S., Harris, G. W., Wostrack, A. & Sansom, C. (2003). *Diffuse X-ray scattering from molecular crystals. Crystallogr. Rev.* **9**, 229–277.

Moss, S. C. (1966). *Local order in solid alloys.* In *Local Atomic Arrangements Studied by X-ray Diffraction*, edited by J. B. Cohen & J. E. Hilliard, ch. 3, pp. 95–122. New York: Gordon and Breach.

Müller, H. (1952). *Die eindimensionale Umwandlung Zinkblende-Wurtzit und die dabei auftretenden Anomalien. Neues Jahrb. Mineral. Abh.* **84**, 43–76.

Neder, R. B., Frey, F. & Schulz, H. (1990). *Diffraction theory for diffuse scattering by correlated microdomains in materials with several atoms per unit cell. Acta Cryst.* A**46**, 792–798.

Nield, V. M. & Keen, D. A. (2001). *Diffuse Neutron Scattering from Crystalline Materials*. Oxford Science Publications.

Nield, V. M., Keen, D. A. & McGreevy, R. L. (1995). *The interpretation of single-crystal diffuse scattering using reverse Monte Carlo modelling. Acta Cryst.* A**51**, 763–771.

Niimura, N. (1986). *Evaluation of data from 1-d PSD used in TOF method. J. Phys. (Paris) Colloq.* **47**(C5), Suppl., 129–136.

Niimura, N., Chataka, T., Ostermann, A., Kurihara, K. & Tanaka, I. (2003). *High resolution neutron protein crystallography. Hydrogen and hydration in proteins. Z. Kristallogr.* **218**, 96–107.

Niimura, N., Ishikawa, Y., Arai, M. & Furusaka, M. (1982). *Applications of position sensitive detectors to structural analysis using pulsed neutron sources. AIP Conference Proceedings*, Vol. 89, *Neutron Scattering*, edited by J. Faber, pp. 11–22. New York: AIP.

Ohshima, K. & Harada, J. (1986). *X-ray diffraction study of short-range ordered structure in a disordered Ag–15 at.% Mg alloy. Acta Cryst.* B**42**, 436–442.

Ohshima, K., Harada, J., Morinaga, M., Georgopoulos, P. & Cohen, J. B. (1986). *Report on a round-robin study of diffuse X-ray scattering. J. Appl. Cryst.* **19**, 188–194.

Ohshima, K. & Moss, S. C. (1983). *X-ray diffraction study of basal-(ab)-plane structure and diffuse scattering from silver atoms in disordered stage-2 $Ag_{0.18}TiS_2$. Acta Cryst.* A**39**, 298–305.

Ohshima, K., Watanabe, D. & Harada, J. (1976). *X-ray diffraction study of short-range order diffuse scattering from disordered Cu–29.8% Pd alloy. Acta Cryst.* A**32**, 883–892.

Overhauser, A. W. (1971). *Observability of charge-density waves by neutron diffraction. Phys. Rev. B*, **3**, 3173–3182.

Pandey, D., Lele, S. & Krishna, P. (1980*a*). *X-ray diffraction from one dimensionally disordered 2H-crystals undergoing solid state transformation to the 6H structure. I. The layer displacement mechanism. Proc. R. Soc. London Ser. A*, **369**, 435–439.

Pandey, D., Lele, S. & Krishna, P. (1980*b*). *X-ray diffraction from one dimensionally disordered 2H-crystals undergoing solid state transformation to the 6H structure. II. The deformation mechanism. Proc. R. Soc. London Ser. A*, **369**, 451–461.

Pandey, D., Lele, S. & Krishna, P. (1980*c*). *X-ray diffraction from one dimensionally disordered 2H-crystals undergoing solid state transformation to the 6H structure. III. Comparison with experimental observations on SiC. Proc. R. Soc. London Ser. A*, **369**, 463–477.

Patterson, A. L. (1959). *Fourier theory.* In *International Tables for X-ray Crystallography*, Vol. II, ch. 2.5. Birmingham: Kynoch Press.

Peisl, J. (1975). *Diffuse X-ray scattering from the displacement field of point defects and defect clusters. J. Appl. Cryst.* **8**, 143–149.

Perez-Mato, J. M., Aroyo, M., Hlinka, J., Quilichini, M. & Currat, R. (1998). *Phonon symmetry selection rules for inelastic neutron scattering. Phys. Rev. Lett.* **81**, 2462–2465.

Petricek, V., Maly, K., Coppens, P., Bu, X., Cisarova, I. & Frost-Jensen, A. (1991). *The description and analysis of composite crystals. Acta Cryst.* A**47**, 210–216.

Pflanz, S. & Moritz, W. (1992). *The domain matrix method: a new calculation scheme for diffraction profiles. Acta Cryst.* A**48**, 716–727.

Pouget, J.-P. (2004). *X-ray diffuse scattering as precursor of incommensurate Peierls transitions in one-dimensional organic transfer conductors. Z. Kristallogr.* **219**, 711–718.

Prandl, W. (1981). *The structure factor of orientational disordered crystals: the case of arbitrary space, site, and molecular point-group. Acta Cryst.* A**37**, 811–818.

Press, W. (1973). *Analysis of orientational disordered structures. I. Method. Acta Cryst.* A**29**, 252–256.

Press, W., Grimm, H. & Hüller, A. (1979). *Analysis of orientational disordered structures. IV. Correlations between orientation and position of a molecule. Acta Cryst.* A**35**, 881–885.

Press, W. & Hüller, A. (1973). *Analysis of orientational disordered structures. II. Examples: Solid CD_4, p-D_2 and $NDBr_4$. Acta Cryst.* A**29**, 257–263.

Proffen, Th. & Neder, R. B. (1999). *DISCUS, a program for diffuse scattering and defect-structure simulations – update. J. Appl. Cryst.* **32**, 838–839.

Proffen, Th. & Welberry, T. R. (1997). *Analysis of diffuse scattering via reverse Monte Carlo technique: a systematic investigation. Acta Cryst.* A**53**, 202–216.

Radons, W., Keller, J. & Geisel, T. (1983). *Dynamical structure factor of a 1-d harmonic liquid: comparison of different approximation methods. Z. Phys. B*, **50**, 289–296.

Rahman, S. H. (1993). *The local domain configuration in partially ordered $AuCu_3$. Acta Cryst.* A**49**, 68–79.

Rietveld, H. M. (1969). *A profile refinement method for nuclear and magnetic structures. J. Appl. Cryst.* **2**, 65–71.

Rosshirt, E., Frey, F. & Boysen, H. (1991). *Disorder diffuse scattering in K-hollandite. Neues Jahrb. Mineral. Abh.* **1**, 101–115.

Rosshirt, E., Frey, F., Boysen, H. & Jagodzinski, H. (1985). *Chain ordering in $E_2PI_{1.6}$ (5,10-diethylphenazinium iodide). Acta Cryst.* B**41**, 66–76.

Sabine, T. M. & Clarke, P. J. (1977). *Powder neutron diffraction – refinement of the total pattern. J. Appl. Cryst.* **10**, 277–280.

Scaringe, P. R. & Ibers, J. A. (1979). *Application of the matrix method to the calculation of diffuse scattering in linearly disordered crystals. Acta Cryst.* A**35**, 803–810.

Schmatz, W. (1973). *X-ray and neutron scattering on disordered crystals.* In *Treatise on Materials Science and Technology*, Vol. 2, edited by H. Hermans, ch. 3.1. New York: Academic Press.

Schmatz, W. (1983). *Neutron scattering studies of lattice defects: static properties of defects.* In *Methods of Experimental Physics, Solid State: Nuclear Physics*, Vol. 21, edited by J. N. Mundy, S. J. Rothman, M. J. Fluss & L. C. Smedskajew, ch. 3.1. New York: Academic Press.

Schulz, H. (1982). *Diffuse X-ray diffraction and its application to materials research.* In *Current Topics in Materials Science*, edited by E. Kaldis, ch. 4. Amsterdam: North-Holland.

Schwartz, L. H. & Cohen, J. B. (1977). *Diffraction from Materials.* New York: Academic Press.

Schweika, W. (1998). *Diffuse scattering and Monte Carlo simulations*. *Springer Tracts in Modern Physics*, Volume 141, edited by G. Höhler. Heidelberg: Springer.

Schweika, W. (2003). *Time-of-flight and vector polarization analysis for diffuse scattering. Physica B*, **335**, 157–163.

Schweika, W. & Böni, P. (2001). *The instrument DNS: Polarization analysis for diffuse neutron scattering. Physica B*, **297**, 155–159.

Sherwood, J. N. (1979). *The Plastically Crystalline State.* New York: John Wiley.

Smaalen, S. van (2004). *An elementary introduction to superspace crystallography. Z. Kristallogr.* **219**, 681–691.

Smaalen, S. van, Lamfers, H.-J. & de Boer, J. L. (1998). *Diffuse scattering and order in quasi-one-dimensional molecular crystals. Phase Transit.* **67**, 277–294.

Sparks, C. J. & Borie, B. (1966). *Methods of analysis for diffuse X-ray scattering modulated by local order and atomic displacements.* In *Atomic Arrangements Studied by X-ray Diffraction*, edited by J. B. Cohen & J. E. Hilliard, ch. 1, pp. 5–50. New York: Gordon and Breach.

Springer, T. (1972). *Quasielastic Neutron Scattering for the Investigation of Diffuse Motions in Solid and Liquids. Springer Tracts in Modern Physics*, Vol. 64. Berlin: Springer.

Steurer, W. & Frey, F. (1998). *Disorder diffuse scattering from quasicrystals. Phase Transit.* **67**, 319–361.

Takaki, Y. & Sakurai, K. (1976). *Intensity of X-ray scattering from one-dimensionally disordered crystal having the multilayer average structure. Acta Cryst.* A**32**, 657–663.

Tibbals, J. E. (1975). *The separation of displacement and substitutional disorder scattering: a correction for structure factor ratio variation. J. Appl. Cryst.* **8**, 111–114.

Turberfield, K. C. (1970). *Time-of-flight diffractometry.* In *Thermal Neutron Diffraction*, edited by B. T. M. Willis. Oxford University Press.

Vainshtein, B. K. (1966). *Diffraction of X-rays by Chain Molecules.* Amsterdam: Elsevier.

Varn, D. P., Canright, G. S. & Crutchfield, J. P. (2002). *Discovering planar disorder in closed-packed structures from X-ray diffraction: Beyond the fault model. Phys. Rev. B*, **66**, 174110–174113.

Warren, B. E. (1941). *X-ray diffraction in random layer lattices. Phys. Rev.* **59**, 693–699.

Warren, B. E. (1969). *X-ray Diffraction.* Reading: Addison-Wesley.

Weber, T. & Bürgi, H.-B. (2002). *Determination and refinement of disordered crystal structures using evolutionary algorithms in combination with Monte Carlo methods. Acta Cryst.* A**58**, 526–540.

Weber, Th. (2005). *Cooperative evolution – a new algorithm for the investigation of disordered structures via Mont Carlo modelling. Z. Kristallogr.* **220**, 1099–1107.

Weber, Th., Boysen, H. & Frey, F. (2000). *Longitudinal positional ordering of n-alkane molecules in urea inclusion compounds. Acta Cryst.* B**56**, 132–141.

Weidner, E., Frey, F. & Hradil, K. (2001). *Transient ordering states in decagonal Al-Co-Ni and Al-Cu-Co-Si phases. Philos. Mag.* **201**, 2375–2389.

Weidner, E., Frey, F., Lei, J.-L., Pedersen, B., Paulmann, C. & Morgenroth, W. (2004). *Disordered quasicrystals: diffuse scattering in decagonal Al-Ni-Fe. J. Appl. Cryst.* **37**, 802–807.

Welberry, T. R. (1983). *Routine recording of diffuse scattering from disordered molecular crystals. J. Appl. Phys.* **16**, 192–197.

Welberry, T. R. (1985). *Diffuse X-ray scattering and models of disorder. Rep. Prog. Phys.* **48**, 1543–1593.

Welberry, T. R. (2004). *Diffuse X-ray Scattering and Models of Disorder*. Oxford University Press.

Welberry, T. R. & Butler, B. D. (1994). *Interpretation of diffuse X-ray scattering via models of disorder. J. Appl. Cryst.* **27**, 205–231.

Welberry, T. R. & Glazer, A. M. (1985). *A comparison of Weissenberg and diffractometer methods for the measurement of diffuse scattering from disordered molecular crystals. Acta Cryst.* A**41**, 394–399.

Welberry, T. R., Goossens, D. J., Haeffner, D. R., Lee, P. L. & Almer, J. (2003). *High-energy diffuse scattering on the 1-ID beamline at the Advanced Photon Source. J. Synchrotron Rad.* **10**, 284–286.

Welberry, T. R. & Proffen, Th. (1998). *Analysis of diffuse scattering from single crystals via the reverse Monte Carlo technique. I. Comparison with direct Monte Carlo. J. Appl. Cryst.* **31**, 309–317.

Welberry, T. R., Proffen, Th. & Bown, M. (1998). *Analysis of single-crystal diffuse X-ray scattering via automatic refinement of a Monte Carlo model. Acta Cryst.* A**54**, 661–674.

Welberry, T. R. & Siripitayananon, J. (1986). *Analysis of the diffuse scattering from disordered molecular crystals: application to 1,4-dibromo-2,5-diethyl-3,6-dimethylbenzene at 295 K. Acta Cryst.* B**42**, 262–272.

Welberry, T. R. & Siripitayananon, J. (1987). *Analysis of the diffuse scattering from disordered molecular crystals: application to 1,3-dibromo-2,5-diethyl-4,6-dimethylbenzene at 295 K. Acta Cryst.* B**43**, 97–106.

Welberry, T. R. & Withers, R. L. (1987). *Optical transforms of disordered systems displaying diffuse intensity loci. J. Appl. Cryst.* **20**, 280–288.

Welberry, T. R. & Withers, R. L. (1990). *Optical transforms of disordered systems containing symmetry-related scattering sites. J. Appl. Cryst.* **23**, 303–314.

Wilke, W. (1983). *General lattice factor of the ideal paracrystal. Acta Cryst.* A**39**, 864–867.

Wilson, A. J. C. (1942). *Imperfections in the structure of cobalt. II. Mathematical treatment of proposed structure. Proc. R. Soc. London Ser. A*, **180**, 277–285.

Wilson, A. J. C. (1949). *X-ray diffraction by random layers: ideal line profiles and determination of structure amplitudes from observed line profiles. Acta Cryst.* **2**, 245–251.

Wilson, A. J. C. (1962). *X-ray Optics*, 2nd ed., chs. V, VI, VIII. London: Methuen.

Windsor, C. G. (1982). *Neutron diffraction performance in pulsed and steady sources.* In *Neutron Scattering. AIP Conference Proceedings*, Vol. 89, edited by J. Faber, pp. 1–10. New York: AIP.

Wolff, P. M. de (1974). *The pseudo-symmetry of modulated crystal structures. Acta Cryst.* A**30**, 777–785.

Wolff, P. M. de, Janssen, T. & Janner, A. (1981). *The superspace groups for incommensurate crystal structures with a one-dimensional modulation. Acta Cryst.* A**37**, 625–636.

Wong, S. F., Gillan, B. E. & Lucas, B. W. (1984). *Single crystal disorder diffuse X-ray scattering from phase II ammonium nitrate, NH_4NO_3. Acta Cryst.* B**40**, 342–346.

Wooster, W. A. (1962). *Diffuse X-ray Reflections from Crystals*, chs. IV, V. Oxford: Clarendon Press.

Wu, T. B., Matsubara, E. & Cohen, J. B. (1983). *New procedures for qualitative studies of diffuse X-ray scattering. J. Appl. Cryst.* **16**, 407–414.

Yessik, M., Werner, S. A. & Sato, H. (1973). *The dependence of the intensities of diffuse peaks on scattering angle in neutron diffraction. Acta Cryst.* A**29**, 372–382.

Young, R. A. (1975). Editor. *International discussion meeting on studies of lattice distortions and local atomic arrangements by X-ray, neutron and electron diffraction. J. Appl. Cryst.* **8**, 79–191.

Zernike, F. & Prins, J. A. (1927). *Die Beugung von Röntgenstrahlen in Flüssigkeiten als Effekt der Molekülanordnung. Z. Phys.* **41**, 184–194.

International Tables for Crystallography (2010). Vol. B, Chapter 4.3, pp. 540–546.

4.3. Diffuse scattering in electron diffraction

By J. M. Cowley† and J. K. Gjønnes

4.3.1. Introduction

The origins of diffuse scattering in electron-diffraction patterns are the same as in the X-ray case: inelastic scattering due to electronic excitations, thermal diffuse scattering (TDS) from atomic motions, scattering from crystal defects or disorder. For diffraction by crystals, the diffuse scattering can formally be described in terms of a nonperiodic deviation $\Delta\varphi$ from the periodic, average crystal potential, $\bar{\varphi}$:

$$\varphi(\mathbf{r}, t) = \bar{\varphi}(\mathbf{r}) + \Delta\varphi(\mathbf{r}, t), \qquad (4.3.1.1)$$

where $\Delta\varphi$ may have a static component from disorder in addition to time-dependent fluctuations of the electron distribution or atomic positions.

In the kinematical case, the diffuse scattering can be treated separately. The intensity I_d as a function of the scattering variable $\mathbf{u}$ ($|\mathbf{u}| = 2\sin\theta/\lambda$) and energy transfer $h\nu$ is then given by the Fourier transform $\mathscr{F}$ of $\Delta\varphi$

$$I(\mathbf{u}, \nu) = |\Delta\Phi(\mathbf{u}\nu)|^2 = |\mathscr{F}\{\Delta\varphi(\mathbf{r}, t)\}|^2 = \mathscr{F}\{P_d(\mathbf{r}, \nu)\} \qquad (4.3.1.2)$$

and may also be written as the Fourier transform of a correlation function P_d representing fluctuations in space and time (see Cowley, 1981). When the energy transfers are small – as with TDS – and hence not measured, the observed intensity corresponds to an integral over ν:

$$I(\mathbf{u}) = I_d(\mathbf{u}) + I_{av}(\mathbf{u})$$
$$I_d(\mathbf{u}) = \int I_d(\mathbf{u}, \nu)\,d\nu = \mathscr{F}\{P_d(\mathbf{r}, \mathbf{0})\}$$

and also

$$I_d(\mathbf{u}) = \langle|\Phi(\mathbf{u})|^2\rangle - |\langle\Phi(\mathbf{u})\rangle|^2, \qquad (4.3.1.3)$$

where the brackets may indicate a time average, an expectation value, or a spatial average over the periodicity of the lattice in the case of static deviations from a periodic structure.

The considerations of TDS and static defects and disorder of Chapters 4.1 and 4.2 thus may be applied directly to electron diffraction in the kinematical approximation when the differences in experimental conditions and diffraction geometry are taken into account.

The most prominent contribution to the diffuse background in electron diffraction, however, is the inelastic scattering at low angles arising mainly from the excitation of outer electrons. This is quite different from the X-ray case where the inelastic ('incoherent') scattering, $S(\mathbf{u})$, goes to zero at small angles and increases to a value proportional to Z for high values of $|\mathbf{u}|$. The difference is due to the Coulomb nature of electron scattering, which leads to the kinematical intensity expression S/u^4, emphasizing the small-angle region. At high angles, the inelastic scattering from an atom is then proportional to Z/u^4, which is considerably less than the corresponding elastic scattering $(Z - f)^2/u^4$ which approaches Z^2/u^4 (Section 2.5.2) (see Fig. 4.3.1.1).

† Deceased.

The kinematical description can be used for electron scattering only when the crystal is very thin (10 nm or less) and composed of light atoms. For heavy atoms such as Au or Pb, crystals of thickness 1 nm or more in principal orientations show strong deviations from kinematical behaviour. With increasing thickness, dynamical scattering effects first modify the sharp Bragg reflections and then have increasingly significant effects on the diffuse scattering. Bragg scattering of the diffuse scattering produces Kikuchi lines and other effects. Multiple diffuse scattering broadens the distribution and smears out detail. As the thickness increases further, the diffuse scattering increases and the Bragg beams are reduced in intensity until there is only a diffuse 'channelling pattern' where the features depend in only a very indirect way on the incident-beam direction or on the sources of the diffuse scattering (Uyeda & Nonoyama, 1968).

The multiple-scattering effects make the quantitative interpretation of diffuse scattering more difficult and complicate the extraction of particular components, *e.g.* disorder scattering. Much of the multiple scattering involves inelastic scattering processes. However, electrons that have lost energy of the order of 1 eV or more can be subtracted experimentally by use of electron energy filters (Krahl *et al.*, 1990; Krivanek *et al.*, 1992) which are commercially available. Measurement can be made also of the complete scattering function $I(\mathbf{u}, \nu)$, but such studies have been rare. Another significant improvement to quantitative measurement of diffuse electron scattering is offered by new recording devices: slow-scan charge-couple-device cameras (Krivanek & Mooney, 1993) and imaging plates (Mori *et al.*, 1990).

There are some advantages in the use of electrons which make it uniquely valuable for particular applications.

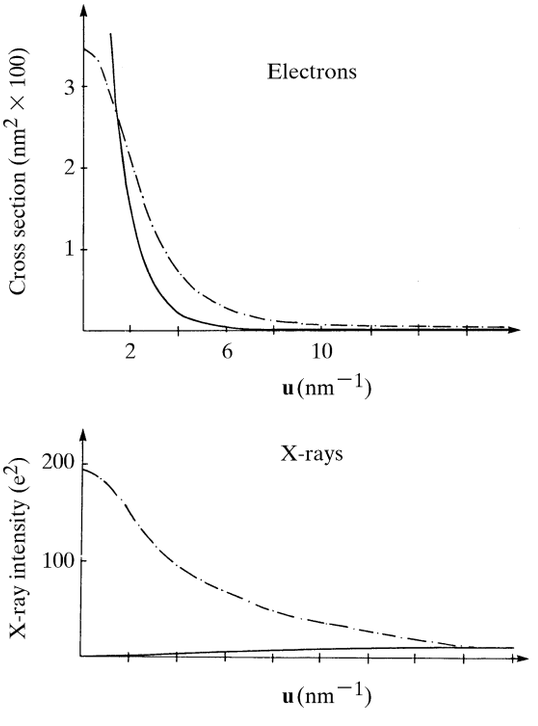


Fig. 4.3.1.1. Comparison between the kinematical inelastic scattering (full line) and elastic scattering (broken) for electrons and X-rays. Values for silicon [Freeman (1960) and *IT* C (2004)].

Copyright © 2010 International Union of Crystallography

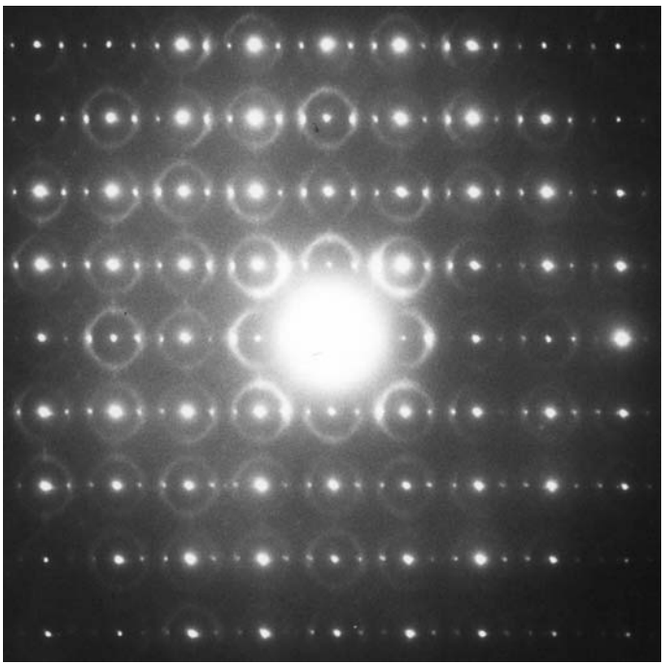


Fig. 4.3.1.2. Electron-diffraction pattern from a disordered crystal of $17Nb_2O_5{\cdot}48WO_3$ close to the [001] orientation of the tetragonal tungsten-bronze-type structure (Iijima & Cowley, 1977).

(1) Diffuse-scattering distributions can be recorded from very small specimen regions, a few nm in diameter and a few nm thick. The diameter of the specimen area may be varied readily up to several μm.

(2) Diffraction information on defects or disorder may be correlated with high-resolution electron-microscope imaging of the same specimen area [see Section 4.3.8 in *IT* C (2004)].

(3) The electron-diffraction pattern approximates to a planar section of reciprocal space, so that complicated configurations of diffuse scattering may be readily visualized (see Fig. 4.3.1.2).

(4) Dynamical effects may be exploited to obtain information about localization of sources of the diffuse scattering within the unit cell.

These experimental and theoretical aspects of electron diffraction have influenced the ways in which it has been applied in studies of diffuse scattering.

In general, we may distinguish three different approaches to the interpretation of diffuse scattering:

(*a*) The crystallographic way, in which the Patterson- or correlation-function representation of the local order is emphasized, *e.g.* by use of short-range-order parameters.

(*b*) The physical model in terms of excitations. These are usually described in reciprocal (momentum) space: phonons, plasmons *etc.*

(*c*) Structure models in direct space. These must be derived by trial or by chemical considerations of bonds, coordinates *etc.*

Owing to the difficulties of separating the different components in the diffuse scattering, most work on diffuse scattering of electrons has followed one or both of the two last approaches, although Patterson-type interpretation, based upon kinematical scattering including some dynamical corrections, has also been tried.

4.3.2. Inelastic scattering

In the kinematical approximation, a general expression which includes inelastic scattering can be written in the form quoted by Van Hove (1954)

$$I(\mathbf{u}, \nu) = \frac{m^3}{2^2h^6}\frac{k}{k_o} \times W(\mathbf{u}) \sum P_{n_o} \sum_{\eta} \sum_{j=1}^{Z} |\langle n_o| \exp\{2\pi i \mathbf{u}\cdot\mathbf{R}_j\}|n\rangle|^2 \times \delta\left(\nu + \frac{E_n - E_{n_o}}{h}\right) \quad (4.3.2.1)$$

for the intensity of scattering as function of energy transfer and momentum transfer from a system of Z identical particles, $\mathbf{R}_j$. Here m and h have their usual meanings; k_o and k, E_{n_o} and E_n are wavevectors and energies before and after the scattering between object states n_o and n; P_{n_o} are weights of the initial states; $W(\mathbf{u})$ is a form factor (squared) for the individual particle.

In equation (4.3.2.1), $\mathbf{u}$ is essentially momentum transfer. When the energy transfer is small ($\Delta E/E \ll \theta$), we can still write $|\mathbf{u}| = 2\sin\theta/\lambda$, then the sum over final states n is readily performed and an expression of the Waller–Hartree type is obtained for the total inelastic scattering as a function of angle:

$$I_{\text{inel}}(\mathbf{u}) \propto \frac{S}{u^4},$$

where

$$S(u) = Z - \sum_{j=1}^{Z} |f_{jj}(u)|^2 - \sum_{j\neq k}^{Z}\sum^{Z} |f_{jk}(u)|^2, \quad (4.3.2.2)$$

and where the one-electron f's for Hartree–Fock orbitals, $f_{jk}(\mathbf{u}) = \langle j|\exp(2\pi i\mathbf{u}\cdot\mathbf{r})|k\rangle$, have been calculated by Freeman (1959, 1960) for atoms up to $Z = 30$. The last sum is over electrons with the same spin only.

The Waller–Hartree formula may be a very good approximation for Compton scattering of X-rays, where most of the scattering occurs at high angles and multiple scattering is no problem. With electrons, it has several deficiencies. It does not take into account the electronic structure of the solid, which is most important at low values of u. It does not include the energy distribution of the scattering. It does not give a finite cross section at zero angle, if u is interpreted as an angle. In order to remedy this, we should go back to equation (4.3.1.2) and decompose $\mathbf{u}$ into two components, one tangential part which is associated with angle in the usual way and one normal component along the beam direction, u_n, which may be related to the excitation energy $\Delta E = E_n - E_{n_o}$ by the expression $u_n = \Delta E_k/2E$. This will introduce a factor $1/(u^2 + u_n^2)$ in the intensity at small angles, often written as $1/(\theta^2 + \theta_E^2)$, with ΔE estimated from ionization energies *etc.* (Strictly speaking, ΔE is not a constant, not even for scattering from one shell. It is a weighted average which will vary with u.)

Calculations beyond this simple adjustment of the Waller–Hartree-type expression are few. Plasmon scattering has been treated on the basis of a nearly free electron model by Ferrel (1957):

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}(\Delta E)\,\mathrm{d}\Omega} = (1/\pi^2 a_H m v^2 N)(-\mathrm{Im}\{1/\varepsilon\})/(\theta^2 + \theta_E^2), \quad (4.3.2.3)$$

where m, v are relativistic mass and velocity of the incident electron, N is the density of the valence electrons and $\varepsilon(\Delta E, \theta)$ their dielectric constant. Upon integration over ΔE:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{E_p}{2\pi a_H m v N}[1/(\theta^2 + \theta_E^2)G(\theta, \theta_c)], \qquad (4.3.2.4)$$

where $G(\theta, \theta_c)$ takes account of the cut-off angle Θ_c. Inner-shell excitations have been studied because of their importance to spectroscopy. The most realistic calculations may be those of Leapman *et al.* (1980) where one-electron wavefunctions are determined for the excited states in order to obtain 'generalized oscillator strengths' which may then be used to modify equation (4.3.1.2).

At high energies and high momentum transfer, the scattering will approach that of free electrons, *i.e.* a maximum at the so-called Bethe ridge, $E = h^2u^2/2m$.

A complete and detailed picture of inelastic scattering of electrons as a function of energy and angle (or scattering variable) is lacking, and may possibly be the least known area of diffraction by solids. It is further complicated by the dynamical scattering, which involves the incident and diffracted electrons and also the ejected atomic electron (see *e.g.* Maslen & Rossouw, 1984*a*,*b*).

4.3.3. Kinematical and pseudo-kinematical scattering

Kinematical expressions for TDS or defect and disorder scattering according to equation (4.3.1.3) can be obtained by inserting the appropriate atomic scattering factors in place of the X-ray scattering factors in Chapter 4.1. The complications introduced by dynamical diffraction are considerable (see Section 4.3.4). In the most general case, a complete specification of the disordered structure may be needed. However, for thin specimens, approximate treatments of the deviations from kinematical scattering may lead to relatively simple forms. Two such cases are treated in this section, both relying on the small-angle nature of electron scattering. The first is based upon the phase-object approximation, which applies to small angles and thin specimens.

The amplitude at the exit surface of a specimen can always be written as a sum of a periodic and a nonperiodic part, and may in analogy with the kinematical case [equation (4.3.1.1)] be written

$$\psi(\mathbf{r}) = \bar{\psi}(\mathbf{r}) + \Delta\psi(\mathbf{r}), \qquad (4.3.3.1)$$

where $\mathbf{r}$ is a vector in two dimensions. The intensities can be separated in the same way [*cf.* equation (4.3.1.3)].

When the phase-object approximation applies [Section 2.5.2.4(*b*)]

$$\begin{aligned}\psi(\mathbf{r}) &= \exp\{-i\sigma\varphi(\mathbf{r})\} \\ &= \exp\{-i\sigma\bar{\varphi}(\mathbf{r})\}[1 - i\sigma\Delta\varphi(\mathbf{r}) - \ldots]. \qquad (4.3.3.2)\end{aligned}$$

Then the Bragg reflections are given by Fourier transform of the periodic part, *viz*:

$$\langle\exp\{-i\sigma\varphi(\mathbf{r})\}\rangle = \exp\{-i\sigma\bar{\varphi}(\mathbf{r})\}\exp\left\{-\tfrac{1}{2}\sigma^2\langle\Delta\varphi^2(\mathbf{r})\rangle\right\}; \qquad (4.3.3.3)$$

note that an absorption function is introduced.

The diffuse scattering derives from

$$-i\sigma\Delta\varphi(\mathbf{r})\exp\{-i\sigma\bar{\varphi}(\mathbf{r})\}, \qquad (4.3.3.4)$$

so that

$$I_d(\mathbf{u}) = \sigma^2|\Delta\Phi(\mathbf{u}) * \Phi_{av}(\mathbf{u})|^2. \qquad (4.3.3.5)$$

Thus, the kinematical diffuse-scattering amplitude is convoluted with the amplitude function for the average structure, *i.e.* the set of sharp Bragg beams. When the direct beam, $\Phi_{av}(0)$, is relatively strong, the kinematical diffuse scattering will be modified to only a limited extent by convolution with the Bragg reflections. To the extent that the diffuse scattering is periodic in reciprocal space, the effect will be to modify the intensity by a slowly varying function. Thus the shapes of local diffuse maxima will not be greatly affected.

The electron-microscope image contrast derived from the diffuse scattering will be obtained by inserting equation (4.3.3.4) in the appropriate intensity expressions of Section 4.3.8 of *IT* C (2004).

Another approach may be used for extended crystal defects in thin films, *e.g.* faults normal or near-normal to the film surface. Often, an average periodic structure may not readily be defined, as in the case of a set of incommensurate stacking faults. Kinematically, the projection of the structure in the simplest case may be described by convoluting the projection of a unit-cell structure with a nonperiodic set of delta functions which constitute a distribution function:

$$\varphi(\mathbf{r}) = \varphi_0(\mathbf{r}) * \sum_n \delta(\mathbf{r} - \mathbf{r}_n) = \varphi_0(\mathbf{r}) * \mathrm{d}(\mathbf{r}). \qquad (4.3.3.6)$$

Then the diffraction-pattern intensity is

$$I(\mathbf{u}) = |\Phi_0(\mathbf{u})|^2|D(\mathbf{u})|^2. \qquad (4.3.3.7)$$

Here, $\Phi_0(\mathbf{u})$ is the scattering amplitude of the unit whereas the function $|D(\mathbf{u})|^2$, where $D(\mathbf{u}) = \mathscr{F}\{d(\mathbf{r})\}$, gives the configuration of spots, streaks or other diffraction maxima corresponding to the faulted structure (see *e.g.* Marks, 1985).

In the projection (column) approximation to dynamical scattering, the wavefunction at the exit surface may be given by an expression identical to (4.3.3.6), but with a wavefunction, $\psi_0(\mathbf{r})$, for the unit in place of the projected potential, $\varphi_0(\mathbf{r})$.

An intensity expression of the same form as (4.3.3.7) then applies, with a dynamical scattering amplitude Ψ_0 for the scattering unit substituted for the kinematical amplitude Φ_0.

$$I(\mathbf{u}) = |\Psi_0(\mathbf{u})|^2|D(\mathbf{u})|^2, \qquad (4.3.3.8)$$

which in the simplest case describes a diffraction pattern with the same features as in the kinematical case. Note that $\Psi_0(\mathbf{u})$ may have different symmetries when the incident beam is tilted away from a zone axis, leading to diffuse streaks *etc.* appearing also in positions where the kinematical diffuse scattering is zero. More complicated cases have been considered by Cowley (1976*a*) who applied this type of analysis to the case of nonperiodic faulting in magnesium fluorogermanate (Cowley, 1976*b*).

4.3.4. Dynamical scattering: Bragg scattering effects

The distribution of diffuse scattering is modified by higher-order terms in essentially two ways: Bragg scattering of the incident and diffuse beams or multiple diffuse scattering, or by a combination.

Theoretical treatment of the Bragg scattering effects in diffuse scattering has been given by many authors, starting with Kainuma's (1955) work on Kikuchi-line contrast (Howie, 1963; Fujimoto & Kainuma, 1963; Gjønnes, 1966; Rez *et al.*, 1977; Maslen & Rossouw, 1984*a*,*b*; Wang, 1995; Allen *et al.*, 1997). Mathematical formalism may vary but the physical pictures and

results are essentially the same. They may be discussed with reference to a Born-series expansion, *i.e.* by introducing the potential φ in the integral equation, as a sum of a periodic and a nonperiodic part [*cf.* equation (4.3.1.1)] and arranging the terms by orders of $\Delta\varphi$.

$$\begin{aligned}\psi &= \psi_0 + G\varphi\psi \\ &= [1 + G\varphi + (G\varphi)^2 + \ldots]\psi_0 \\ &= [1 + G\bar{\varphi} + (G\bar{\varphi})^2 + \ldots]\psi_0 \\ &\quad + [1 + G\bar{\varphi} + (G\bar{\varphi})^2 + \ldots] \\ &\quad \times G(\Delta\varphi)[1 + G\bar{\varphi} + (G\bar{\varphi})^2 + \ldots]\psi_0 \\ &\quad + \text{higher-order terms.} \qquad (4.3.4.1)\end{aligned}$$

Some of the higher-order terms contributing to the Bragg scattering can be included by adding the essentially imaginary term $\langle\Delta\varphi G(\Delta\varphi)\rangle$ to the static potential $\bar{\varphi}$.

Theoretical treatments have mostly been limited to the first-order diffuse scattering. With the usual approximation to forward scattering, the expression for the amplitude of diffuse scattering in a direction $\mathbf{k}_0 + \mathbf{u} + \mathbf{g}$ can be written as

$$\begin{aligned}\psi(\mathbf{u}+\mathbf{g}) = \sum_g \sum_f \int_0^z & S_{hg}(\mathbf{k}_0 + \mathbf{u}, z - z_1) \\ & \times \Delta\Phi(\mathbf{u} + \mathbf{g} - \mathbf{f}) S_{f0}(\mathbf{k}_0, z_1)\, \mathrm{d}z_1 \qquad (4.3.4.2)\end{aligned}$$

and read (from right to left): S_{f0}, Bragg scattering of the incident beam above the level z_1; $\Delta\Phi$, diffuse scattering within a thin layer $\mathrm{d}z_1$ through the Fourier components $\Delta\Phi$ of the nonperiodic potential $\Delta\Phi$; S_{hg}, Bragg scattering between diffuse beams in the lower part of the crystal. It is commonly assumed that diffuse scattering at different levels can be treated as independent (Gjønnes, 1966), then the intensity expression becomes

$$\begin{aligned}I(\mathbf{u}+\mathbf{g}) = \sum_h \sum_{h'} \sum_f \sum_{f'} \int_0^z & S_{gh}(2) S^*_{gh'}(2) \\ & \times \langle \Delta\Phi(\mathbf{u} + \mathbf{h} - \mathbf{f}) \Delta\Phi^*(\mathbf{u} + \mathbf{h}' - \mathbf{f}') \rangle \\ & \times S_{f0}(1) S^*_{f'0}(1)\, \mathrm{d}z_1, \qquad (4.3.4.3)\end{aligned}$$

where (1) and (2) refer to the regions above and below the diffuse-scattering layer. This expression can be manipulated further, *e.g.* by introducing Bloch-wave expansion of the scattering matrices, *viz*

$$\begin{aligned}I(\mathbf{u}+\mathbf{g}) = \sum_{hh'} \sum_{ff'} \sum_{jj'} \sum_{ii'} & C^j_g(2) C^{j'*}_g(2) C^{j*}_h(2) C^{j'}_{h'}(2) \\ & \times \frac{\exp[i(\gamma^i - \gamma^{i'})z] - \exp[i(\gamma^j - \gamma^{j'})z]}{\gamma^i - \gamma^{i'} - \gamma^j + \gamma^{j'}} \\ & \times \langle f(\mathbf{u} + \mathbf{h} - \mathbf{f}) f^*(\mathbf{u} + \mathbf{h}' - \mathbf{f}') \rangle \\ & \times C^{i*}_f(1) C^{i'}_{f'}(1) C^i_0(1) C^{i*}_0(1), \qquad (4.3.4.4)\end{aligned}$$

which may be interpreted as scattering by $\Delta\varphi$ between Bloch waves belonging to the same branch (intraband scattering) or different branches (interband scattering). Another alternative is to evaluate the scattering matrices by multislice calculations (Section 4.3.5).

Expressions such as (4.3.4.2) contain a large number of terms. Unless very detailed calculations relating to a precisely defined model are to be carried out, attention should be focused on the most important terms.

In Kikuchi-line contrast, the scattering in the upper part of the crystal is usually not considered and frequently the angular variation of the $\Phi(\mathbf{u})$ is also neglected. In diffraction contrast from small-angle inelastic scattering, it may be sufficient to consider the intraband terms [$i = i'$, $j = j'$ in (4.3.4.3)].

In studies of diffuse-scattering distribution, the factor $\langle\Phi(\mathbf{u}+\mathbf{h})\Phi(\mathbf{u}+\mathbf{h}')\rangle$ will produce two types of terms: Those with $\mathbf{h} = \mathbf{h}'$ result only in a redistribution of intensity between corresponding points in the Brillouin zones, with the same total intensity. Those with $\mathbf{h} \neq \mathbf{h}'$ lead to enhancement or reduction of the total diffuse intensity and hence absorption from the Bragg beam and enhanced/reduced intensity of secondary radiation, *i.e.* anomalous absorption and channelling effects. They arise through interference between different Fourier components of the diffuse scattering and carry information about position of the sources of diffuse scattering, referred to the projected unit cell. This is exploited in channelling experiments, where beam direction is used to determine atom reaction (Taftø & Spence, 1982; Taftø & Lehmpfuhl, 1982).

Gjønnes & Høier (1971) expressed this information in terms of the Fourier transform R of the generalized or Kikuchi-line form factor;

$$\langle\Delta\Phi(\mathbf{u})\Delta\Phi^*(\mathbf{u}+\mathbf{h})\rangle\mathbf{u}^2(\mathbf{u}+\mathbf{h})^2 = Q(\mathbf{u}+\mathbf{h}) = \mathscr{F}\{R(\mathbf{r},\mathbf{g})\} \qquad (4.3.4.5)$$

includes information both about correlations between sources of diffuse scattering and about their position in the projected unit cell. It is seen that $Q(\mathbf{u}, 0)$ represents the kinematical intensity, hence $\int R(\mathbf{r},\mathbf{g})\,\mathrm{d}\mathbf{g}$ is the Patterson function. The integral of $Q(\mathbf{u},\mathbf{h})$ in the plane gives the anomalous absorption (Yoshioka, 1957) which is related to the distribution $R(\mathbf{r}, 0)$ of scattering centres across the unit cell.

The scattering factor $\langle\Delta\Phi(\mathbf{u})\Delta\Phi^*(\mathbf{u}-\mathbf{h})\rangle$ can be calculated for different modes. For one-electron excitations as an extension of the Waller–Hartree expression (Gjønnes, 1962; Whelan, 1965):

$$\begin{aligned}\langle\Delta\Phi(\mathbf{u})\Delta\Phi^*(\mathbf{u}-\mathbf{h})\rangle = & \sum_i \frac{f_{ii}(\mathbf{h}) - f_{ii}(\mathbf{u}) f_{ii}(|\mathbf{h}-\mathbf{u}|)}{u^2(\mathbf{h}-\mathbf{u})^2} \\ & - \sum_i \sum_{i\neq j} \frac{f_{ij}(\mathbf{u}) f_{ij}(|\mathbf{h}-\mathbf{u}|)}{u^2(\mathbf{h}-\mathbf{u})^2}, \qquad (4.3.4.6)\end{aligned}$$

where f_{ij} are the one-electron amplitudes (Freeman, 1959).

A similar expression for scattering by phonons is obtained in terms of the scattering factors $G_j(\mathbf{u},\mathbf{g})$ for the branch j, wavevector $\mathbf{g}$ and a polarization vector $\mathbf{l}_{j,q}$ (see Chapter 4.1):

$$\langle\Delta\Phi(\mathbf{u})\Delta\Phi^*(\mathbf{u}-\mathbf{h})\rangle = G_j(\mathbf{u},\mathbf{q})G^*_j(\mathbf{u}-\mathbf{h},\mathbf{q}), \qquad (4.3.4.7)$$

independent phonons being assumed.

For scattering from substitutional order in a binary alloy with ordering on one site only, we obtain simply

$$\Delta\Phi(\mathbf{u})\Delta\Phi^*(\mathbf{u}-\mathbf{h}) = |\Delta\varphi(\mathbf{u})|^2 \frac{f_A(|\mathbf{u}-\mathbf{h}|) - f_B(|\mathbf{u}-\mathbf{h}|)}{f_A(\mathbf{u}) - f_B(\mathbf{u})}, \qquad (4.3.4.8)$$

where $f_{A,B}$ are atomic scattering factors. It is seen that $Q(\mathbf{u})$ then does not contain any new information; the location of the site involved in the ordering is known.

When several sites are involved in the ordering, the dynamical scattering factor becomes less trivial, since scattering factors for

the different ordering parameters (for different sites) will include a factor $\exp(2\pi i \mathbf{r}_m \cdot \mathbf{h})$ (see Andersson *et al.*, 1974).

From the above expressions, it is found that the Bragg scattering will affect diffuse scattering from different sources differently: Diffuse scattering from substitutional order will usually be enhanced at low and intermediate angles, whereas scattering from thermal and electronic fluctuations will be reduced at low angles and enhanced at higher angles. This may be used to study substitutional order and displacement order (size effect) separately (Andersson, 1979).

The use of such expressions for quantitative or semi-quantitative interpretation raises several problems. The Bragg scattering effects occur in all diffuse components, in particular the inelastic scattering, which thus may no longer be represented by a smooth, monotonic background. It is best to eliminate this experimentally. When this cannot be done, the experiment should be arranged so as to minimize Kikuchi-line excess/deficient terms, by aligning the incident beam along a not too dense zone. In this way, one may optimize the diffuse-scattering information and minimize the dynamical corrections, which then are used partly as guides to conditions, partly as refinement in calculations.

The multiple scattering of the background remains as the most serious problem. Theoretical expressions for multiple scattering in the absence of Bragg scattering have been available for some time (Moliere, 1948), as a sum of convolution integrals

$$I(\mathbf{u}) = [(\mu t)I_1(\mathbf{u}) + (1/2)(\mu t)^2 I_2(\mathbf{u}) + \ldots] \exp(-\mu t), \quad (4.3.4.9)$$

where $I_2(\mathbf{u}) = I_1(\mathbf{u}) * I_1(\mathbf{u}) \ldots$ *etc.*, and $I_1(\mathbf{u})$ is normalized.

A complete description of multiple scattering in the presence of Bragg scattering should include Bragg scattering between diffuse scattering at all levels z_1, z_2, *etc.* This quickly becomes unwieldy. Fortunately, the experimental patterns seem to indicate that this is not necessary: The Kikuchi-line contrast does not appear to be very sensitive to the exact Bragg condition of the incident beam. Høier (1973) therefore introduced Bragg scattering only in the last part of the crystal, *i.e.* between the level z_n and the final thickness z for n-times scattering. He thus obtained the formula:

$$I(\mathbf{u}+\mathbf{h}) = \sum_j |C_h^j|^2 \left\{ A_1^j \sum_g \sum_{g \neq g'} F_1(\mathbf{u}, \mathbf{g}, \mathbf{g}') C_g^j C_{g'}^{j*} + \sum_n \sum_g A_{ng}^j F_n(\mathbf{u}, \mathbf{g}) |C_g^j|^2 \right\}, \quad (4.3.4.10)$$

where F_n are normalized scattering factors for nth-order multiple diffuse scattering and A_n^j are multiple-scattering coefficients which include absorption.

When the thickness is increased, the variation of $F_n(\mathbf{u}, \mathbf{g})$ with angle becomes slower, and an expression for intensity of the channelling pattern is obtained (Gjønnes & Taftø, 1976):

$$I(\mathbf{u}+\mathbf{h}) = \sum_j \sum_g \sum_n |C_h^j|^2 |C_g^j|^2 A_n^j = \sum_j |C_h^j|^2 \sum_n A_n^j \rightarrow |C_h^j|^2 / \mu^j(\mathbf{u}). \quad (4.3.4.11)$$

Another approach is the use of a modified diffusion equation (Ohtsuki *et al.*, 1976).

These expressions seem to reproduce the development of the general background with thickness over a wide range of thicknesses. It may thus appear that the contribution to the diffuse background from known sources can be treated adequately – and that such a procedure must be included together with adequate filtering of the inelastic component in order to improve the quantitative interpretation of diffuse scattering.

4.3.5. Multislice calculations for diffraction and imaging

The description of dynamical diffraction in terms of the progression of a wave through successive thin slices of a crystal (Chapter 5.2) forms the basis for the multislice method for the calculation of electron-diffraction patterns and electron-microscope images [see Section 4.3.6.1 in *IT* C (2004)]. This method can be applied directly to the calculations of diffuse scattering in electron diffraction due to thermal motion and positional disorder and for calculating the images of defects in crystals.

It is essentially an amplitude calculation based on the formulation of equation (4.3.4.1) [or (4.3.4.2)] for first-order diffuse scattering. The Bragg scattering in the first part of the crystal is calculated using a standard multislice method for the set of beams **h**. In the nth slice of the crystal, a diffuse-scattering amplitude $\Psi_d(\mathbf{u})$ is convoluted with the incident set of Bragg beams. For each **u**, propagation of the set of beams $\mathbf{u}+\mathbf{h}$ is then calculated through the remaining slices of the crystal. The intensities for the exit wave at the set of points $\mathbf{u}+\mathbf{h}$ are then calculated by adding either amplitudes or intensities. Amplitudes are added if there is correlation between the defects in successive slices. Intensities are added if there are no such correlations. The process is repeated for all **u** values to obtain a complete mapping of the diffuse scattering.

Calculations have been made in this way, for example, for short-range order in alloys [Fisher (1969); see also Cowley (1981) ch. 17] and also for TDS on the assumption of both correlated and uncorrelated atomic motions (Doyle, 1969). The effects of the correlations were shown to be small.

This computing method is not practical for electron-microscope images in which individual defects are to be imaged. The perturbations of the exit wavefunction due to individual defects (vacancies, replaced atoms, displaced atoms) or small groups of defects may then be calculated with arbitrary accuracy by use of the 'periodic continuation' form of the multislice computer programs in which an artificial, large, superlattice unit cell is assumed [Section 4.3.6.1 in *IT* C (2004)]. The corresponding images and microdiffraction patterns from the individual defects or clusters may then be calculated (Fields & Cowley, 1978). A more recent discussion of the image calculations, particularly in relation to thermal diffuse scattering, is given by Cowley (1988).

In order to calculate the diffuse-scattering distributions from disordered systems or from a crystal with atoms in thermal motion by the multislice method with periodic continuation, it would be necessary to calculate for a number of different defect configurations sufficiently large to provide an adequate representation of the statistics of the disordered system. However, it has been shown by Cowley & Fields (1979) that, if the single-diffuse-scattering approximation is made, the perturbations of the exit wave due to individual defects are characteristic of the defect type and of the slice number and may be added, so that a considerable simplification of the computing process is possible. Methods for calculating diffuse scattering in electron-diffraction patterns using the multislice approach are described by Tanaka & Cowley (1987) and Cowley (1989). Loane *et al.* (1991) introduced the concept of 'frozen phonons' for multislice calculations of thermal scattering.

4.3.6. Qualitative interpretation of diffuse scattering of electrons

Quantitative interpretation of the intensity of diffuse scattering by calculation of *e.g.* short-range-order parameters has been the exception. Most studies have been directed to qualitative features

and their variation with composition, treatment *etc.* Many features in the scattering which pass unrecognized in extensive X-ray or neutron investigations will be observed readily with electrons, frequently inviting other ways of interpretation.

Most such studies have been concerned with substitutional disorder, but the extensive investigations of thermal streaks by Honjo and co-workers should be mentioned (Honjo *et al.*, 1964). Diffuse spots and streaks from disorder have been observed from a wide range of substances. The most frequent may be streaks due to planar faults, one of the most common objects studied by electron microscopy. Diffraction patterns are usually sufficient to determine the orientation and the fault vector; the positions and distribution of faults are more easily seen by dark-field microscopy, whereas the detailed atomic arrangement is best studied by high-resolution imaging of the structure [Section 4.3.8 in *IT* C (2004)].

This combination of diffraction and different imaging techniques cannot be applied in the same way to the study of the essentially three-dimensional substitutional local order. Considerable effort has therefore been made to interpret the details of diffuse scattering, leaving the determination of the short-range-order (SRO) parameters usually to X-ray or neutron studies.

Frequently, characteristic shapes or splitting of the diffuse spots from *e.g.* binary alloys are observed. They reflect order extending over many atomic distances, and have been assumed to arise from forces other than the near-neighbour pair forces invoked in the theory of local order. A relationship between the diffuse-scattering distribution and the Fourier transform of the effective atom-pair-interaction potential is given by the ordering theory of Clapp & Moss (1968). An interpretation in terms of long-range forces carried by the conduction electrons was proposed by Krivoglaz (1969). Extensive studies of alloy systems (Ohshima & Watanabe, 1973) show that the separations, m, observed in split diffuse spots from many alloys follow the predicted variation with the electron/atom ratio e/a:

$$m = \left[\frac{12}{\pi}(e/a)\right]^{1/3} t - \sqrt{2},$$

where m is measured along the [110] direction in units of $2a^*$ and t is a truncation factor for the Fermi surface.

A similarity between the location of diffuse maxima and the shape of the Fermi surface has been noted also for other structures, notably some defect rock-salt-type structures. Although this may offer a clue to the forces involved in the ordering, it entails no description of the local structure. Several attempts have been made to formulate principles for building the disordered structure, from small ordered domains embedded in less ordered regions (Hashimoto, 1974), by a network of antiphase boundaries, or by building the structure from clusters with the average composition and coordination (De Ridder *et al.*, 1976). Evidence for such models may be sought by computer simulations, in the details of the SRO scattering as seen in electron diffraction, or in images.

The cluster model is most directly tied to the location of diffuse scattering, noting that a relation between order parameters derived from clusters consistent with the ordered state can be used to predict the position of diffuse scattering in the form of surfaces in reciprocal space, *e.g.* the relation $\cos h + \cos k + \cos l = 0$ for ordering of octahedral clusters in the rock-salt-type structure (Sauvage & Parthé, 1974).

Some of the models imply local fluctuations in order which may be observable either by diffraction from very small regions or by imaging. Microdiffraction studies (Tanaka & Cowley, 1985) do indeed show that spots from 1–1.5 nm regions in disordered $LiFeO_2$ appear on the locus of diffuse maxima observed in diffraction from larger areas.

Imaging of local variations in the SRO structure has been pursued with different techniques (De Ridder *et al.*, 1976; Tanaka & Cowley, 1985; De Meulenaare *et al.*, 1998), *viz*: dark field using diffuse spots only; bright field with the central spot *plus* diffuse spots; lattice image. With domains of about 3 nm or more, high-resolution images seem to give clear indication of their presence and form. For smaller ordered regions, the interpretation becomes increasingly complex: Since the domains will then usually not extend through the thickness of the foil, they cannot be imaged separately. Since image-contrast calculations essentially demand complete specification of the local structure, a model beyond the statistical description must be constructed in order to be compared with observations. On the other hand, these models of the local structure should be consistent with the statistics derived from diffraction patterns collected from a larger volume.

References

Allen, L. J., Josefsson, T. W., Lehmpfuhl, G. & Uchida, Y. (1997). *Modeling thermal diffuse scattering in electron diffraction involving higher-order Laue zones*. *Acta Cryst.* A**53**, 421–425.

Andersson, B. (1979). *Electron diffraction study of diffuse scattering due to atomic displacements in disordered vanadium monoxide. Acta Cryst.* A**35**, 718–727.

Andersson, B., Gjønnes, J. & Taftø, J. (1974). *Interpretation of short range order scattering of electrons; application to ordering of defects in vanadium monoxide. Acta Cryst.* A**30**, 216–224.

Clapp, P. C. & Moss, S. C. (1968). *Correlation functions of disordered binary alloys. II. Phys. Rev.* **171**, 754–763.

Cowley, J. M. (1976*a*). *Diffraction by crystals with planar faults. I. General theory. Acta Cryst.* A**32**, 83–87.

Cowley, J. M. (1976*b*). *Diffraction by crystals with planar faults. II. Magnesium fluorogermanate. Acta Cryst.* A**32**, 88–91.

Cowley, J. M. (1981). *Diffraction Physics*, 2nd ed. Amsterdam: North-Holland.

Cowley, J. M. (1988). *Electron microscopy of crystals with time-dependent perturbations. Acta Cryst.* A**44**, 847–853.

Cowley, J. M. (1989). *Multislice methods for surface diffraction and inelastic scattering.* In *Computer Simulation of Electron Microscope Diffraction and Images,* edited by W. Krakow & M. O'Keefe, pp. 1–12. Worrendale: The Minerals, Metals and Materials Society.

Cowley, J. M. & Fields, P. M. (1979). *Dynamical theory for electron scattering from crystal defects and disorder. Acta Cryst.* A**35**, 28–37.

De Meulenaare, P., Rodewald, M. & Van Tendeloo, G. (1998). *Anisotropic cluster model for the short-range order in Cu_{1-x}–Pd_x-type alloys*. *Phys. Rev. B*, **57**, 11132–11140.

De Ridder, R., Van Tendeloo, G. & Amelinckx, S. (1976). *A cluster model for the transition from the short-range order to the long-range order state in f.c.c. based binary systems and its study by means of electron diffraction. Acta Cryst.* A**32**, 216–224.

Doyle, P. A. (1969). *Dynamical calculation of thermal diffuse scattering. Acta Cryst.* A**25**, 569–577.

Ferrel, R. A. (1957). *Characteristic energy loss of electrons passing through metal foils. Phys. Rev.* **107**, 450–462.

Fields, P. M. & Cowley, J. M. (1978). *Computed electron microscope images of atomic defects in f.c.c. metals. Acta Cryst.* A**34**, 103–112.

Fisher, P. M. J. (1969). *The development and application of an n-beam dynamic methodology in electron diffraction.* PhD thesis, University of Melbourne.

Freeman, A. J. (1959). *Compton scattering of X-rays from non-spherical charge distributions. Phys. Rev.* **113**, 169–175.

Freeman, A. J. (1960). *X-ray incoherent scattering functions for non-spherical charge distributions. Acta Cryst.* **12**, 929–936.

Fujimoto, F. & Kainuma, Y. (1963). *Inelastic scattering of fast electrons by thin crystals. J. Phys. Soc. Jpn*, **18**, 1792–1804.

Gjønnes, J. (1962). *Inelastic interaction in dynamic electron scattering. J. Phys. Soc. Jpn*, **17**, Suppl. BII, 137–139.

Gjønnes, J. (1966). *The influence of Bragg scattering on inelastic and other forms of diffuse scattering of electrons. Acta Cryst.* **20**, 240–249.

Gjønnes, J. & Høier, R. (1971). *Structure information from anomalous absorption effects in diffuse scattering of electrons. Acta Cryst.* A**27**, 166–174.

Gjønnes, J. & Taftø, J. (1976). *Bloch wave treatment of electron channelling. Nucl. Instrum. Methods*, **133**, 141–148.
Hashimoto, S. (1974). *Correlative microdomain model for short range ordered alloy structures. I. Diffraction theory. Acta Cryst.* A**30**, 792–798.
Høier, R. (1973). *Multiple scattering and dynamical effects in diffuse electron scattering. Acta Cryst.* A**29**, 663–672.
Honjo, G., Kodera, S. & Kitamura, N. (1964). *Diffuse streak diffraction patterns from single crystals. I. General discussion and aspects of electron diffraction diffuse streak patterns. J. Phys. Soc. Jpn*, **19**, 351–369.
Howie, A. (1963). *Inelastic scattering of electrons by crystals. I. The theory of small angle inelastic scattering. Proc. Phys. Soc. A*, **271**, 268–287.
Iijima, S. & Cowley, J. M. (1977). *Study of ordering using high resolution electron microscopy. J. Phys. (Paris)*, **38**, Suppl. C7, 21–30.
International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.
Kainuma, Y. (1955). *The theory of Kikuchi patterns. Acta Cryst.* **8**, 247–257.
Krahl, D., Pätzold, H. & Swoboda, M. (1990). *An aberration-minimized imaging energy filter of simple design. Proceedings of the 12th International Conference on Electron Microscopy*, Vol. 2, pp. 60–61.
Krivanek, O. L., Gubbens, A. J., Dellby, N. & Meyer, C. E. (1992). *Design and first applications of a post-column imaging filter. Micros. Microanal. Microstruct. (France)*, **3**, 187–199.
Krivanek, O. L. & Mooney, P. E. (1993). *Applications of slow-scan CCD cameras in transmission electron microscopy. Ultramicroscopy*, **49**, 95–108.
Krivoglaz, M. A. (1969). *Theory of X-ray and Thermal Neutron Scattering by Real Crystals.* New York: Plenum.
Leapman, R. D., Rez, P. & Mayers, D. F. (1980). *L and M shell generalized oscillator strength and ionization cross sections for fast electron collisions. J. Chem. Phys.* **72**, 1232–1243.
Loane, R. F., Xu, P. R. & Silcox, J. (1991). *Thermal vibrations in convergent-beam electron diffraction. Acta Cryst.* A**47**, 267–278.
Marks, L. D. (1985). *Image localization. Ultramicroscopy*, **18**, 33–38.
Maslen, W. V. & Rossouw, C. J. (1984*a*). *Implications of (e, 2e) scattering electron diffraction in crystals: I. Philos. Mag. A*, **49**, 735–742.
Maslen, W. V. & Rossouw, C. J. (1984*b*). *Implications of (e, 2e) scattering electron diffraction in crystals: II. Philos. Mag. A*, **49**, 743–757.
Moliere, G. (1948). *Theory of scattering of fast charged particles: plural and multiple scattering. Z. Naturforsch. Teil A*, **3**, 78–97.
Mori, M., Oikawa, T. & Harada, Y. (1990). *Development of the imaging plate for the transmission electron microscope and its characteristics. J. Electron Microsc. (Jpn)*, **19**, 433–436.
Ohshima, K. & Watanabe, D. (1973). *Electron diffraction study of short-range-order diffuse scattering from disordered Cu–Pd and Cu–Pt alloys. Acta Cryst.* A**29**, 520–525.
Ohtsuki, Y. H., Kitagaku, M., Waho, T. & Omura, T. (1976). *Dechannelling theory with the Fourier–Planck equation and a modified diffusion coefficient. Nucl. Instrum Methods*, **132**, 149–151.
Rez, P., Humphreys, C. J. & Whelan, M. J. (1977). *The distribution of intensity in electron diffraction patterns due to phonon scattering. Philos. Mag.* **35**, 81–96.
Sauvage, M. & Parthé, E. (1974). *Prediction of diffuse intensity surfaces in short-range-ordered ternary derivative structures based on ZnS, NaCl, CsCl and other structures. Acta Cryst.* A**30**, 239–246.
Taftø, J. & Lehmpfuhl, G. (1982). *Direction dependence in electron energy loss spectroscopy from single crystals. Ultramicroscopy*, **7**, 287–294.
Taftø, J. & Spence, J. C. H. (1982). *Atomic site determination using channelling effect in electron induced X-ray emission. Ultramicroscopy*, **9**, 243–247.
Tanaka, N. & Cowley, J. M. (1985). *High resolution electron microscopy of disordered lithium ferrites. Ultramicroscopy*, **17**, 365–377.
Tanaka, N. & Cowley, J. M. (1987). *Electron microscope imaging of short range order in disordered alloys. Acta Cryst.* A**43**, 337–346.
Uyeda, R. & Nonoyama, M. (1968). *The observation of thick specimens by high voltage electron microscopy. Jpn. J. Appl. Phys.* **1**, 200–208.
Van Hove, L. (1954). *Correlations in space and time and Born approximation scattering in systems of interacting particles. Phys. Rev.* **95**, 249–262.
Wang, Z. L. (1995). *Elastic and Inelastic Scattering in Electron Diffraction and Imaging*. New York: Plenum Press.
Whelan, M. (1965). *Inelastic scattering of fast electrons by crystals. I. Interband excitations. J. Appl. Phys.* **36**, 2099–2110.
Yoshioka, H. (1957). *Effect of inelastic waves on electron diffraction. J. Phys. Soc. Jpn*, **12**, 618–628.

International Tables for Crystallography (2010). Vol. B, Chapter 4.4, pp. 547–566.

4.4. Scattering from mesomorphic structures

By P. S. Pershan

4.4.1. Introduction

The term mesomorphic is derived from the prefix 'meso-', which is defined in the dictionary as 'a word element meaning middle', and the term '-morphic', which is defined as 'an adjective termination corresponding to morph or form'. Thus, mesomorphic order implies some 'form', or order, that is 'in the middle', or intermediate between that of liquids and crystals. The name liquid crystalline was coined by researchers who found it to be more descriptive, and the two are used synonymously. It follows that a mesomorphic, or liquid-crystalline, phase must have more symmetry than any one of the 230 space groups that characterize crystals.

A major source of confusion in the early liquid-crystal literature was concerned with the fact that many of the molecules that form liquid crystals also form true three-dimensional crystals with diffraction patterns that are only subtly different from those of other liquid-crystalline phases. Since most of the original mesomorphic phase identifications were performed using a 'miscibility' procedure, which depends on optically observed changes in textures accompanying variation in the sample's chemical composition, it is not surprising that some three-dimensional crystalline phases were mistakenly identified as mesomorphic. Phases were identified as being either the same as, or different from, phases that were previously observed (Liebert, 1978; Gray & Goodby, 1984), and although many of the workers were very clever in deducing the microscopic structure responsible for the microscopic textures, the phases were labelled in the order of discovery as smectic-A, smectic-B *etc.* without any attempt to develop a systematic nomenclature that would reflect the underlying order. Although different groups did not always assign the same letters to the same phases, the problem is now resolved and the assignments used in this article are commonly accepted (Gray & Goodby, 1984).

Fig. 4.4.1.1 illustrates the way in which increasing order can be assigned to the series of mesomorphic phases in three dimensions listed in Table 4.4.1.1. Although the phases in this series are the most thoroughly documented mesomorphic phases, there are others not included in the table which we will discuss below.

The progression from the completely symmetric isotropic liquid through the mesomorphic phases into the crystalline phases can be described in terms of three separate types of order. The first, or the molecular orientational order, describes the fact that the molecules have some preferential orientation analogous to the spin orientational order of ferromagnetic materials. In the present case, the molecular quantity that is oriented is a symmetric second-rank tensor, like the moment of inertia or the electric polarizability, rather than a magnetic moment. This is the only type of long-range order in the nematic phase and as a consequence its physical properties are those of an anisotropic fluid; this is the origin of the name liquid crystal. Fig. 4.4.1.2(*a*) is a schematic illustration of the nematic order if it is assumed that the molecules can be represented by oblong ellipses. The average orientation of the ellipses is aligned; however, there is no long-range order in the relative positions of the ellipses. Nematic phases are also observed for disc-shaped molecules and for clusters of molecules that form micelles. These all share the common properties of being optically anisotropic and fluid-like, without any long-range positional order.

The second type of order is referred to as bond orientational order. Consider, for example, the fact that for dense packing of spheres on a flat surface most of the spheres will have six neighbouring spheres distributed approximately hexagonally around it. If a perfect two-dimensional triangular lattice of indefinite size were constructed of these spheres, each hexagon on the lattice would be oriented in the same way. Within the last few years, we have come to recognize that this type of order, in which the hexagons are everywhere parallel to one another, is possible even when there is no lattice. This type of order is referred to as bond orientational order, and bond orientational order in the absence of a lattice is the essential property defining the hexatic phases (Halperin & Nelson, 1978; Nelson & Halperin, 1979; Young, 1979; Birgeneau & Litster, 1978).

The third type of order is the positional order of an indefinite lattice of the type that defines the 230 space groups of conven-

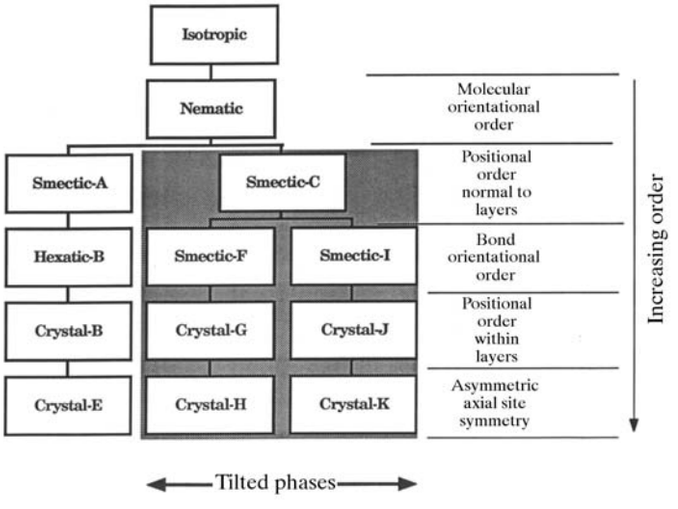


Fig. 4.4.1.1. Illustration of the progression of order throughout the sequence of mesomorphic phases that are based on 'rod-like' molecules. The shaded section indicates phases in which the molecules are tilted with respect to the smectic layers.

Table 4.4.1.1. *Some of the symmetry properties of the series of three-dimensional phases described in Fig. 4.4.1.1*

The terms LRO and SRO imply long-range or short-range order, respectively, and QLRO refers to 'quasi-long-range order' as explained in the text.

Phase	Molecular orientation order within layer	Bond orientation order	Positional order	
			Normal to layer	Within layer
Smectic-A (SmA)	SRO	SRO	SRO	SRO
Smectic-C (SmC)	LRO	LRO†	SRO	SRO
Hexatic-B	LRO†	LRO	QLRO	SRO
Smectic-F (SmF)	LRO	LRO	QLRO	SRO
Smectic-I (SmI)	LRO	LRO	QLRO	SRO
Crystalline-B (CrB)	LRO	LRO	LRO	LRO
Crystalline-G (CrG)	LRO	LRO	LRO	LRO
Crystalline-J (CrJ)	LRO	LRO	LRO	LRO
Crystalline-E (CrE)	LRO	LRO	LRO	LRO
Crystalline-H (CrH)	LRO	LRO	LRO	LRO
Crystalline-K (CrK)	LRO	LRO	LRO	LRO

† Theoretically, the existence of LRO in the molecular orientation, or tilt, implies that there must be some LRO in the bond orientation and *vice versa*.

Copyright © 2010 International Union of Crystallography

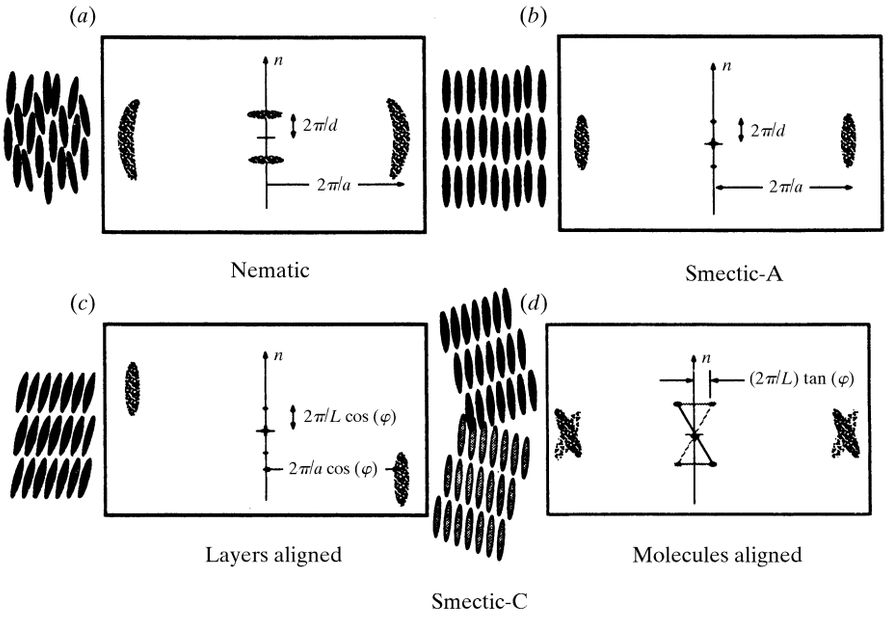


Fig. 4.4.1.2. Schematic illustration of the real-space molecular order and the scattering cross sections in reciprocal space for the: (*a*) nematic; (*b*) smectic-A; and (*c*), (*d*) smectic-C phases. The scattering cross sections are enclosed in the boxes. Part (*c*) indicates the smectic-C phase for an oriented monodomain and (*d*) indicates a polydomain smectic-C structure in which the molecular axes are aligned.

tional crystals. In view of the fact that some of the mesomorphic phases have a layered structure, it is convenient to separate the positional order into the positional order along the layer normal and perpendicular to it, or within the layers.

Two of the symmetries listed in Tables 4.4.1.1 and 4.4.1.2 are short-range order (SRO), implying that the order is only correlated over a finite distance such as for a simple liquid, and long-range order (LRO) as in either the spin orientation of a ferromagnet or the positional order of a three-dimensional crystal. The third type of symmetry, 'quasi-long-range order' (QLRO), will be explained below. In any case, the progressive increase in symmetry from the isotropic liquid to the crystalline phases for this series of mesomorphic phases is illustrated in Fig. 4.4.1.1. One objective of this chapter is to describe the reciprocal-space structure of the phases listed in the tables and the phase transitions between them.

Finally, in most of the crystalline phases that we wish to discuss, the molecules have considerable amounts of rotational disorder. For example, one series of molecules that form mesomorphic phases consists of long thin molecules which might be described as 'blade shaped'. Although the cross section of these molecules is quite anisotropic, the site symmetry of the molecule is often symmetric, as though the molecule is rotating freely about its long axis. On cooling, many of the mesomorphic systems undergo transitions to the phases, listed at the bottom of Fig. 4.4.1.1, for which the site symmetry is anisotropic as though some of the rotational motions about the molecular axis have been frozen out. A similar type of transition, in which rotational motions are frozen out, occurs on cooling systems such as succinonitrile ($NCCH_2CH_2CN$) that form optically isotropic 'plastic crystals' (Springer, 1977).

Table 4.4.1.2. *The symmetry properties of the two-dimensional hexatic and crystalline phases*

Phase	Molecular orientation order within layer	Bond orientation order	Positional order within layer
Smectic-A (SmA)	SRO	SRO	SRO
Smectic-C (SmC)	QLRO	QLRO	SRO
Hexatic-B	QLRO	QLRO	SRO
Smectic-F (SmF)	QLRO	QLRO	SRO
Smectic-I (SmI)	QLRO	QLRO	SRO
Crystalline-B (CrB)	LRO	LRO	QLRO
Crystalline-G (CrG)	LRO	LRO	QLRO
Crystalline-J (CrJ)	LRO	LRO	QLRO
Crystalline-E (CrE)	LRO	LRO	QLRO
Crystalline-H (CrH)	LRO	LRO	QLRO
Crystalline-K (CrK)	LRO	LRO	QLRO

There are two broad classes of liquid-crystalline systems, the thermotropic and the lyotropic, and, since the former are much better understood, this chapter will emphasize results on thermotropic systems (Liebert, 1978). The historical difference between these two, and also the origin of their names, is that the lyotropic are always mixtures, or solutions, of unlike molecules in which one is a normal, or nonmesogenic, liquid. Solutions of soap and water are prototypical examples of lyotropics, and their mesomorphic phases appear as a function of either concentration or temperature. In contrast, the thermotropic systems are usually formed from a single chemical component, and the mesomorphic phases appear primarily as a function of temperature changes. The molecular distinction between the two is that one of the molecules in the lyotropic solution always has a hydrophilic part, often called the 'head group', and one or more hydrophobic alkane chains called 'tails'. These molecules will often form mesomorphic phases as single-component or neat systems; however, the general belief is that in solution with either water or oil most of the phases are the result of competition between the hydrophilic and hydrophobic interactions, as well as other factors such as packing and steric constraints (Pershan, 1979; Safran & Clark, 1987). To the extent that molecules that form thermotropic liquid-crystalline phases have hydrophilic and hydrophobic parts, the disparity in the affinity of these parts for either water or oil is much less and most of these molecules are relatively insoluble in water. These molecules are called thermotropic because their phase transformations are primarily studied only as a function of temperature. This is not to say that there are not numerous examples of interesting studies of the concentration dependence of phase diagrams involving mixtures of thermotropic liquid crystals.

Fig. 4.4.1.3 displays some common examples of molecules that form lyotropic and thermotropic phases. In spite of the above remarks, it is interesting to observe that different parts of typical

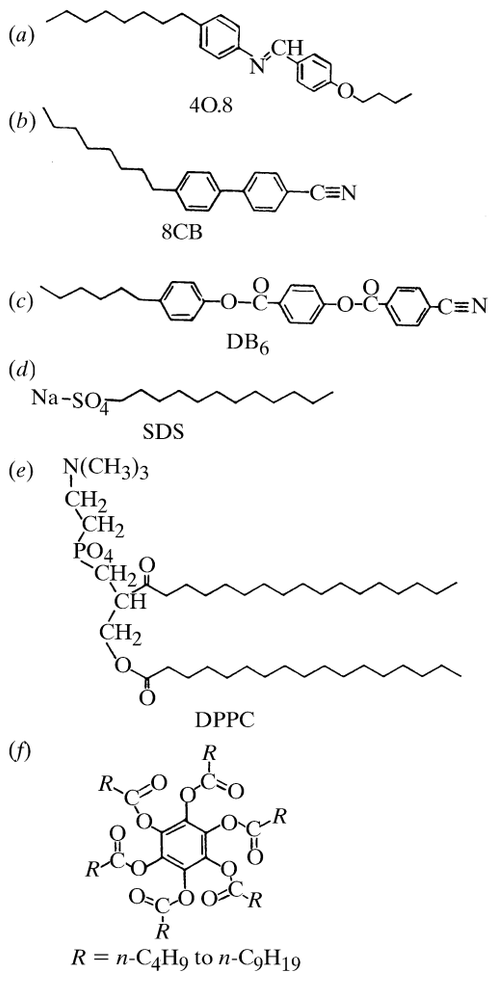


Fig. 4.4.1.3. Chemical formulae for some of the molecules that form thermotropic liquid crystals: (*a*) *N*-[4-(*n*-butyloxy)benzylidene]-4-*n*-octylaniline (4O.8), (*b*) 4′-*n*-octylbiphenyl-4-carbonitrile (8CB), (*c*) 4-hexylphenyl 4-(4-cyanobenzoyloxy)benzoate (DB_6); lyotropic liquid crystals: (*d*) sodium dodecyl sulfate, (*e*) 1,2-dipalmitoyl-L-phosphatidylcholine (DPPC); and a discotic liquid crystal: (*f*) benzenehexayl hexa-*n*-alkanoates.

thermotropic molecules do have some of the same features as the lyotropic molecules. For example, although the rod-like thermotropic molecules always have an alkane chain at one or both ends of a more rigid section, the chain lengths are rarely as long as those of the lyotropic molecules, and although the solubility of the parts of the thermotropic molecules, when separated, are not as disparate as those of the lyotropic molecules, they are definitely different. We suspect that this may account for the subtler features of the phase transformations between the mesomorphic phases to be discussed below. On the other hand, the inhomogeneity of the molecule is probably not important for the nematic phase.

4.4.2. The nematic phase

The nematic phase is a fluid for which the molecules have long-range orientational order. The phase as well as its molecular origin can be most simply illustrated by treating the molecules as long thin rods. The orientation of each molecule can be described by a symmetric second-rank tensor $s_{i,j} = (n_i n_j - \delta_{i,j}/3)$, where $\mathbf{n}$ is a unit vector along the axis of the rod (De Gennes, 1974). For disc-like molecules, such as that shown in Fig. 4.4.1.3(*f*), or for micellar nematic phases, $\mathbf{n}$ is along the principal symmetry axis of either the molecule or the micelle (Lawson & Flautt, 1967). Since physical quantities such as the molecular polarizability, or the moment of inertia, transform as symmetric second-rank tensors, either one of these could be used as specific representations of the molecular orientational order. The macroscopic order, however, is given by the statistical average $S_{i,j} = \langle s_{i,j} \rangle = S(\langle n_i \rangle \langle n_j \rangle - \delta_{i,j}/3)$, where $\langle \mathbf{n} \rangle$ is a unit vector along the macroscopic symmetry axis and S is the order parameter of the nematic phase.

The microscopic origin of the phase can be understood in terms of steric constraints that occur on filling space with highly asymmetric objects such as long rods or flat discs. Maximizing the density requires some degree of short-range orientational order, and theoretical arguments can be invoked to demonstrate long-range order. Onsager presented quantitative arguments of this type to explain the nematic order observed in concentrated solutions of the long thin rods of tobacco mosaic viruses (Onsager, 1949; Lee & Meyer, 1986), and qualitatively similar ideas explain the nematic order for the shorter thermotropic molecules (Maier & Saupe, 1958, 1959).

The existence of nematic order can also be understood in terms of a phenomenological mean-field theory (De Gennes, 1969*b*, 1971; Fan & Stephen, 1970). If the free-energy difference ΔF between the isotropic and nematic phases can be expressed as an analytic function of the nematic order parameter $S_{i,j}$, one can expand $\Delta F(S_{i,j})$ as a power series in which the successive terms all transform as the identity representation of the point group of the isotropic phase, *i.e.* as scalars. The most general form is given by

$$\Delta F(S_{i,j}) = \frac{A}{2}\sum_{ij} S_{ij}S_{ji} + \frac{B}{3}\sum_{ijk} S_{ij}S_{jk}S_{ki} + \frac{D}{4}\left|\sum_{ij} S_{ij}S_{ji}\right|^2 + \frac{D'}{4}\sum_{ijkl} S_{ij}S_{jk}S_{kl}S_{li}. \qquad (4.4.2.1)$$

The usual mean-field treatment assumes that the coefficient of the leading term is of the form $A = a(T - T^*)$, where T is the absolute temperature and T^* is the temperature at which $A = 0$. Taking a, D and $D' > 0$, one can show that for either positive or negative values of B, but for sufficiently large T, the minimum value of $\Delta F = 0$ occurs for $S_{i,j} = 0$, corresponding to the isotropic phase. For $T < T^*$, ΔF can be minimized, at some negative value, for a nonzero $S_{i,j}$ corresponding to nematic order. The details of how this is derived for a tensorial order parameter can be found in the literature (De Gennes, 1974); however, the basic idea can be understood by treating $S_{i,j}$ as a scalar. If we write

$$\Delta F = \tfrac{1}{2}AS^2 + \tfrac{1}{3}BS^3 + \tfrac{1}{4}DS^4 = \left\{\frac{A}{2} - \frac{B^2}{9D}\right\}S^2 + \frac{D}{4}\left(S + \frac{2B}{3D}\right)^2 S^2 \qquad (4.4.2.2)$$

and if T_{NI} is defined by the condition $A = a(T_{NI} - T^*) = 2B^2/9D$, then $\Delta F = 0$ for both $S = 0$ and $S = -2B/3D$. This value for T_{NI} marks the transition temperature from the isotropic phase, when $A > 2B^2/9D$ and the only minimum is at $S = 0$ with $\Delta F = 0$, to the nematic case when $A < 2B^2/9D$ and the absolute minimum with $\Delta F < 0$ is slightly shifted from $S = -2B/3D$. The symmetry properties of second-rank tensors imply that there will usually be a nonvanishing value for B, and this implies that the transition from the isotropic to nematic transition will be first order with a discontinuous jump in the nematic order parameter $S_{i,j}$. Although most nematic systems are uniaxial, biaxial nematic order is theoretically possible (Freiser, 1971; Alben, 1973; Lubensky, 1987) and it has been observed in certain lyotropic nematic liquid crystals (Neto *et al.*,

1985; Hendrikx *et al.*, 1986; Yu & Saupe, 1980) and in one thermotropic system (Malthête *et al.*, 1986).

The X-ray scattering cross section of an oriented monodomain sample of the nematic phase with rod-like molecules usually exhibits a diffuse spot like that illustrated in Fig. 4.4.1.2(*a*), where the maximum of the cross section is along the average molecular axis $\langle \mathbf{n} \rangle$ at a value of $|\mathbf{q}| \approx 2\pi/d$, where $d \approx 20.0$ to 40.0 Å is of the order of the molecular length L. This is a precursor to the smectic-A order that develops at lower temperatures for many materials. In addition, there is a diffuse ring along the directions normal to $\langle \mathbf{n} \rangle$ at $|\mathbf{q}| \approx 2\pi/a$, where $a \approx 4.0$ Å is comparable to the average radius of the molecule. In some nematic systems, the near-neighbour correlations favour antiparallel alignment and molecular centres tend to form pairs such that the peak of the scattering cross section can actually have values anywhere in the range from $2\pi/L$ to $2\pi/2L$. There are also other cases where there are two diffuse peaks, corresponding to both $|\mathbf{q}_1| \approx 2\pi/L$ and $|\mathbf{q}_2| \approx |\mathbf{q}_1|/2$ which are precursors of a richer smectic-A morphology (Prost & Barois, 1983; Prost, 1984; Sigaud *et al.*, 1979; Wang & Lubensky, 1984; Hardouin *et al.*, 1983; Chan, Pershan *et al.*, 1985). In some cases, $|\mathbf{q}_2| \not\approx \frac{1}{2}|\mathbf{q}_1|$ and competition between the order parameters at incommensurate wavevectors gives rise to modulated phases. For the moment, we will restrict the discussion to those systems for which the order parameter is characterized by a single wavevector.

On cooling, many nematic systems undergo a second-order phase transition to a smectic-A phase and as the temperature approaches the nematic to smectic-A transition the widths of these diffuse peaks become infinitesimally small. De Gennes (1972) demonstrated that this phenomenon could be understood by analogy with the transitions from either normal fluidity to superfluidity in liquid helium or normal conductivity to superconductivity in metals. Since the electron density of the smectic-A phase is (quasi-)periodic in one dimension, he represented it by the form

$$\rho(\mathbf{r}) = \langle \rho \rangle + \mathrm{Re}\{\Psi \exp[i(2\pi/d)z]\},$$

where d is the thickness of the smectic layers lying in the xy plane. The complex quantity $\psi = |\psi| \exp(i\varphi)$ is similar to the superfluid wavefunction except that in this analogy the amplitude $|\psi|$ describes the electron-density variations normal to the smectic layers, and the phase φ describes the position of the layers along the z axis. De Gennes proposed a mean-field theory for the transition in which the free-energy difference between the nematic and smectic-A phase $\Delta F(\Psi)$ was represented by

$$\Delta F(\psi) = \frac{A}{2}|\psi|^2 + \frac{D}{4}|\psi|^4 + \frac{E}{2}\left[\left|\left\{\frac{\partial}{\partial z} - i\left(\frac{2\pi}{d}\right)\right\}\psi\right|^2 + \left(\frac{\partial \psi}{\partial x}\right)^2 + \left(\frac{\partial \psi}{\partial y}\right)^2\right]. \quad (4.4.2.3)$$

This mean-field theory differs from the one for the isotropic to nematic transition in that the symmetry for the latter allowed a term that was cubic in the order parameter, while no such term is allowed for the nematic to smectic-A transition. In both cases, however, the coefficient of the leading term is taken to have the form $a(T - T^*)$. If $D > 0$, without the cubic term the free energy has only one minimum when $T > T^*$ at $|\psi| = 0$, and two equivalent minima at $|\psi| = \{a(T^* - T)/D\}^{0.5}$ for $T < T^*$. On the basis of this free energy, the nematic to smectic-A transition can be second order with a transition temperature $T_{\mathrm{NA}} = T^*$ and an order parameter that varies as the square root of $(T_{\mathrm{NA}} - T)$. There are conditions that we will not discuss in detail when D can be negative. In that case, the nematic to smectic-A transition will be first order (McMillan, 1972, 1973*a*,*b*,*c*). McMillan pointed out that, by allowing coupling between the smectic and nematic order parameters, a more general free energy can be developed in which D is negative. McMillan's prediction that for systems in which the difference $T_{\mathrm{IN}} - T_{\mathrm{NA}}$ is small the nematic to smectic-A transition will be first order is supported by experiment (Ocko, Birgeneau & Litster, 1986; Ocko *et al.*, 1984; Thoen *et al.*, 1984). Although the mean-field theory is not quantitatively accurate, it does explain the principal qualitative features of the nematic to smectic-A transition.

The differential scattering cross section for X-rays can be expressed in terms of the Fourier transform of the density–density correlation function $\langle \rho(\mathbf{r})\rho(0) \rangle$. The expectation value is calculated from the thermal average of the order parameter that is obtained from the free-energy density $\Delta F(\psi)$. If one takes the transform

$$\Psi(\mathbf{Q}) \equiv \frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{r} \exp[i(\mathbf{Q}\cdot\mathbf{r})]\rho(\mathbf{r}), \quad (4.4.2.4)$$

the free-energy density in reciprocal space has the form

$$\Delta F(\psi) = \frac{A}{2}|\psi|^2 + \frac{D}{4}|\psi|^4 + \frac{E}{2}\{[Q_z - (2\pi/d)]^2 + Q_x^2 + Q_y^2\}|\psi|^2 \quad (4.4.2.5)$$

and one can show that for $T > T_{\mathrm{NA}}$ the cross section obtained from the above form for the free energy is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \approx \frac{\sigma_0}{A + E\{[Q_z - (2\pi/d)]^2 + Q_x^2 + Q_y^2\}}, \quad (4.4.2.6)$$

where the term in $|\psi|^4$ has been neglected. The mean-field theory predicts that the peak intensity should vary as $\sigma_0/A \approx 1/(T - T_{\mathrm{NA}})$ and that the half width of the peak in any direction should vary as $(A/E)^{1/2} \approx (T - T_{\mathrm{NA}})^{1/2}$. The physical interpretation of the half width is that the smectic fluctuations in the nematic phase are correlated over lengths $\xi = (E/A)^{1/2} \sim (T - T_{\mathrm{NA}})^{-1/2}$.

One of the major shortcomings of all mean-field theories is that they do not take into account the difference between the average value of the order parameter $\langle \psi \rangle$ and the instantaneous value $\psi = \langle \psi \rangle + \delta\psi$, where $\delta\psi$ represents the thermal fluctuations (Ma, 1976). The usual effect expected from theories for this type of critical phenomenon is a 'renormalization' of the various terms in the free energy such that the temperature dependence of correlation length has the form $\xi(t) \propto t^{-\nu}$, where $t \equiv (T - T^*)/T^*$, $T^* = T_{\mathrm{NA}}$ is the second-order transition temperature, and ν is expected to have some universal value that is generally not equal to 0.5. One of the major unsolved problems of the nematic to smectic-A phase transition is that the width along the scattering vector $\mathbf{q}$ varies as $1/\xi_\parallel \propto t^{\nu_\parallel}$ with a temperature dependence different from that of the width perpendicular to $\mathbf{q}$, $1/\xi_\perp \propto t^{\nu_\perp}$; also, neither $\nu_\parallel$ nor $\nu_\perp$ have the expected universal values (Lubensky, 1983; Nelson & Toner, 1981).

The correlation lengths are measured by fitting the differential scattering cross sections to the empirical form:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{\sigma}{1 + (\mathbf{Q}_z - |\mathbf{q}|)^2\xi_\parallel^2 + \mathbf{Q}_\perp^2\xi_\perp^2 + c(\mathbf{Q}_\perp^2\xi_\perp^2)^2}. \quad (4.4.2.7)$$

The amplitude $\sigma \propto t^{-\gamma}$, where the measured values of γ are empirically found to be very close to the measured values for the sum $\nu_\parallel + \nu_\perp$. Most of the systems that have been measured to date have values for $\nu_\parallel > 0.66 > \nu_\perp$ and $\nu_\parallel - \nu_\perp \approx 0.1$ to 0.2. Table 4.4.2.1 lists sources of the observed values for γ, $\nu_\parallel$ and $\nu_\perp$. The theoretical and experimental studies of this pretransition effect account for a sizeable fraction of all of the liquid-crystal research in the last 15 or 20 years, and as of this writing the explanation for these two different temperature dependences remains one of the major unresolved theoretical questions in equilibrium statistical physics.

It is very likely that the origin of the problem is the QLRO in the position of the smectic layers. Lubensky attempted to deal with this by introducing a gauge transformation in such a way that the thermal fluctuations of the transformed order parameter did not have the logarithmic divergence. While this approach has been informative, it has not yet yielded an agreed-upon understanding. Experimentally, the effect of the phase can be studied in systems where there are two competing order parameters with wavevectors that are at $\mathbf{q}_2$ and $\mathbf{q}_1 \approx 2\mathbf{q}_2$ (Sigaud *et al.*, 1979; Hardouin *et al.*, 1983; Prost & Barois, 1983; Wang & Lubensky, 1984; Chan, Pershan *et al.*, 1985). On cooling, mixtures of 4-hexylphenyl 4-(4-cyanobenzoyloxy)benzoate (DB_6) and *N*,*N*′-(1,4-phenylenedimethylene)bis(4-butylaniline) (also known as terephthal-bis-butylaniline, TBBA) first undergo a second-order transition from the nematic to a phase that is designated as smectic-A_1. The various smectic-A and smectic-C morphologies will be described in more detail in the following section; however, the smectic-A_1 phase is characterized by a single peak at $q_1 = 2\pi/d$ owing to a one-dimensional density wave with wavelength d of the order of the molecular length L. In addition, however, there are thermal fluctuations of a second-order parameter with a period of $2L$ that give rise to a diffuse peak at $q_2 = \pi/L$. On further cooling, this system undergoes a second second-order transition to a smectic-A_2 phase with QLRO at $q_2 \approx \pi/L$, with a second harmonic that is exactly at $q = 2q_2 \approx 2\pi/L$. The critical scattering on approaching this transition is similar to that of the nematic to smectic-A_1, except that the pre-existing density wave at $q_1 = 2\pi/L$ quenches the phase fluctuations of the order parameter at the subharmonic $q_2 = \pi/L$. The measured values of $\nu_\parallel = \nu_\perp \approx 0.74$ (Chan, Pershan *et al.*, 1985) agree with those expected from the appropriate theory (Huse, 1985). A mean-field theory that describes this effect is discussed in Section 4.4.3.2 below.

It is interesting to note that even those systems for which the nematic to smectic-A transition is first order show some pretransitional lengthening of the correlation lengths $\xi_\parallel$ and $\xi_\perp$. In these cases, the apparent T^* at which the correlation lengths would diverge is lower than T_{NA} and the divergence is truncated by the first-order transition (Ocko *et al.*, 1984).

Table 4.4.2.1. *Summary of critical exponents from X-ray scattering studies of the nematic to smectic-A phase transition*

Molecule	γ	$\nu_\parallel$	$\nu_\perp$	Reference
4O.7	1.46	0.78	0.65	(*a*)
$\bar{8}$S5	1.53	0.83	0.68	(*b*), (*g*)
CBOOA	1.30	0.70	0.62	(*c*), (*d*)
4O.8	1.31	0.70	0.57	(*e*)
8OCB	1.32	0.71	0.58	(*d*), (*f*)
$\bar{9}$S5	1.31	0.71	0.57	(*b*), (*g*)
8CB	1.26	0.67	0.51	(*h*), (*i*)
10̄S5	1.10	0.61	0.51	(*b*), (*g*)
9CB	1.10	0.57	0.39	(*g*), (*j*)

References: (*a*) Garland *et al.* (1983); (*b*) Brisbin *et al.* (1979); (*c*) Djurek *et al.* (1974); (*d*) Litster *et al.* (1979); (*e*) Birgeneau *et al.* (1981); (*f*) Kasting *et al.* (1980); (*g*) Ocko *et al.* (1984); (*h*) Thoen *et al.* (1982); (*i*) Davidov *et al.* (1979); (*j*) Thoen *et al.* (1984).

4.4.3. Smectic-A and smectic-C phases

4.4.3.1. Homogeneous smectic-A and smectic-C phases

In the smectic-A and smectic-C phases, the molecules organize themselves into layers, and from a naive point of view one might describe them as forming a one-dimensional periodic lattice in which the individual layers are two-dimensional liquids. In the smectic-A phase, the average molecular axis $\langle \mathbf{n} \rangle$ is normal to the smectic layers while for the smectic-C it makes a finite angle. It follows from this that the smectic-C phase has lower symmetry than the smectic-A, and the phase transition from the smectic-A to smectic-C can be considered as the ordering of a two-component order parameter, *i.e.* the two components of the projection of the molecular axis on the smectic layers (De Gennes, 1973). Alternatively, Chen & Lubensky (1976) have developed a mean-field theory in which the transition is described by a free-energy density of the Lifshitz form. This will be described in more detail below; however, it corresponds to replacing equation (4.4.2.5) for the free energy $\Delta F(\psi)$ by an expression for which the minimum is obtained when the wavevector $\mathbf{q}$, of the order parameter $\psi \propto \exp[i\mathbf{q} \cdot \mathbf{r}]$, tilts away from the molecular axis.

The X-ray cross section for the prototypical aligned monodomain smectic-A sample is shown in Fig. 4.4.1.2(*b*). It consists of a single sharp spot along the molecular axis at $|\mathbf{q}|$ somewhere between $2\pi/2L$ and $2\pi/L$ that reflects the QLRO along the layer normal, and a diffuse ring in the perpendicular direction at $|\mathbf{q}| \approx 2\pi/a$ that reflects the SRO within the layer. The scattering cross section for an aligned smectic-C phase is similar to that of the smectic-A except that the molecular tilt alters the intensity distribution of the diffuse ring. This is illustrated in Fig. 4.4.1.2(*c*) for a monodomain sample. Fig. 4.4.1.2(*d*) illustrates the scattering pattern for a polydomain smectic-C sample in which the molecular axis remains fixed, but where the smectic layers are randomly distributed azimuthally around the molecular axis.

The naivety of describing these as periodic stacks of two-dimensional liquids derives from the fact that the sharp spot along the molecular axis has a distinct temperature-dependent shape indicative of QLRO that distinguishes it from the Bragg peaks due to true LRO in conventional three-dimensional crystals. Landau and Peierls discussed this effect for the case of two-dimensional crystals (Landau, 1965; Peierls, 1934) and Caillé (1972) extended the argument to the mesomorphic systems.

The usual treatment of thermal vibrations in three-dimensional crystals estimates the Debye–Waller factor by integrating the thermal expectation value for the mean-square amplitude over reciprocal space (Kittel, 1963):

$$W \simeq \frac{k_B T}{c^3} \int_0^{k_D} \frac{k^{(d-1)}}{k^2} \mathrm{d}k, \qquad (4.4.3.1)$$

where c is the sound velocity, $\omega_D \equiv ck_D$ is the Debye frequency and $d = 3$ for three-dimensional crystals. In this case, the integral converges and the only effect is to reduce the integrated intensity of the Bragg peak by a factor proportional to $\exp(-2W)$. For two-dimensional crystals $d = 1$, and the integral, of the form of $\mathrm{d}k/k$, obtains a logarithmic divergence at the lower limit (Fleming *et al.*, 1980). A more precise treatment of thermal vibrations, necessitated by this divergence, is to calculate the relative phase of X-rays scattered from two points in the sample a distance $|\mathbf{r}|$ apart. The appropriate integral that replaces the Debye–Waller integral is

$$\langle [u(\mathbf{r}) - u(0)]^2 \rangle \simeq \frac{k_B T}{c^3} \int \sin^2(\mathbf{k} \cdot \mathbf{r}) \frac{\mathrm{d}k}{k} \mathrm{d}\{\cos(\mathbf{k} \cdot \mathbf{r})\} \qquad (4.4.3.2)$$

and the divergence due to the lower limit is cut off by the fact that $\sin^2(\mathbf{k}\cdot\mathbf{r})$ vanishes as $k \to 0$. More complete analysis obtains $\langle[u(\mathbf{r}) - u(0)]^2\rangle \simeq (k_B T/c^2)\ln(|\mathbf{r}|/a)$, where $a \approx$ atomic size. If this is exponentiated, as for the Debye–Waller factor, the density–density correlation function can be shown to have the form $\langle\rho(\mathbf{r})\rho(0)\rangle \simeq |\mathbf{r}/a|^{-\eta}$, where $\eta \simeq |\mathbf{q}|^2(k_B T/c^2)$ and $|\mathbf{q}| \simeq 2\pi/a$. In place of the usual periodic density–density correlation function of three-dimensional crystals, the periodic correlations of two-dimensional crystals decay as some power of the distance. This type of positional order, in which the correlations decay as some power of the distance, is the quasi-long-range order (QLRO) that appears in Tables 4.4.1.1 and 4.4.1.2. It is distinguished from true long-range order (LRO) where the correlations continue indefinitely, and short-range order (SRO) where the positional correlations decay exponentially as in either a simple fluid or a nematic liquid crystal.

The usual prediction of Bragg scattering for three-dimensional crystals is obtained from the Fourier transform of the three-dimensional density–density correlation function. Since the correlation function is made up of periodic and random parts, it follows that the scattering cross section is made up of a δ function at the Bragg condition superposed on a background of thermal diffuse scattering from the random part. In principle, these two types of scattering can be separated empirically by using a high-resolution spectrometer that integrates all of the δ-function Bragg peak, but only a small part of the thermal diffuse scattering. Since the two-dimensional lattice is not strictly periodic, there is no formal way to separate the periodic and random parts, and the Fourier transform for the algebraic correlation function obtains a cross section that is described by an algebraic singularity of the form $|\mathbf{Q} - \mathbf{q}|^{\eta-2}$ (Gunther *et al.*, 1980). In 1972, Caillé (Caillé, 1972) presented an argument that the X-ray scattering line shape for the one-dimensional periodicity of the smectic-A system in three dimensions has an algebraic singularity that is analogous to the line shapes from two-dimensional crystals.

In three-dimensional crystals, both the longitudinal and the shear sound waves satisfy linear dispersion relations of the form $\omega = ck$. In simple liquids, and also for nematic liquid crystals, only the longitudinal sound wave has such a linear dispersion relation. Shear sound waves are overdamped and the decay rate $1/\tau$ is given by the imaginary part of a dispersion relation of the form $\omega = i(\eta/\rho)k^2$, where η is a viscosity coefficient and ρ is the liquid density. The intermediate order of the smectic-A mesomorphic phase, between the three-dimensional crystal and the nematic, results in one of the modes for shear sound waves having the curious dispersion relation $\omega^2 = c^2 k_\perp^2 k_z^2/(k_\perp^2 + k_z^2)$, where $k_\perp$ and k_z are the magnitudes of the components of the acoustic wavevector perpendicular and parallel to $\langle\mathbf{n}\rangle$, respectively (De Gennes, 1969*a*; Martin *et al.*, 1972). More detailed analysis, including terms of higher order in $k_\perp^2$, obtains the equivalent of the Debye–Waller factor for the smectic-A as

$$W \simeq k_B T \int_0^{k_D} \frac{k_\perp \mathrm{d}k_\perp \mathrm{d}k_z}{Bk_z^2 + Kk_\perp^4}, \qquad (4.4.3.3)$$

where B and K are smectic elastic constants, $k_\perp^2 = k_x^2 + k_y^2$, and k_D is the Debye wavevector. On substitution of $u^2 = (K/B)k_\perp^2 + k_z^2$, the integral can be manipulated into the form $\int \mathrm{d}u/u$, which diverges logarithmically at the lower limit in exactly the same way as the integral for the Debye–Waller factor of the two-dimensional crystal. The result is that the smectic-A phase has a sharp peak, described by an algebraic cusp, at the place in reciprocal space where one would expect a true δ-function Bragg cross section from a truly periodic one-dimensional lattice. In fact, the lattice is not truly periodic and the smectic-A system has only QLRO along the direction $\langle\mathbf{n}\rangle$.

X-ray scattering experiments to test this idea were carried out on one thermotropic smectic-A system, but the results, while consistent with the theory, were not adequate to provide an unambiguous proof of the algebraic cusp (Als-Nielsen *et al.*, 1980). One of the principal difficulties was due to the fact that, when thermotropic samples are oriented in an external magnetic field in the higher-temperature nematic phase and then gradually cooled through the nematic to smectic-A phase transition, the smectic-A samples usually have mosaic spreads of the order of a fraction of a degree and this is not sufficient for detailed line-shape studies near to the peak. A second difficulty is that, in most of the thermotropic smectic-A phases that have been studied to date, only the lowest-order peak is observed. It is not clear whether this is due to a large Debye–Waller-type effect or whether the form factor for the smectic-A layer falls off this rapidly. Nevertheless, since the factor η in the exponent of the cusp $|\mathbf{Q} - \mathbf{q}|^{\eta-2}$ depends quadratically on the magnitude of the reciprocal vector $|\mathbf{q}|$, the shape of the cusp for the different orders would constitute a severe test of the theory.

Fortunately, it is common to observe multiple orders for lyotropic smectic-A systems and such an experiment, carried out on the lyotropic smectic-A system formed from a quaternary mixture of sodium dodecyl sulfate, pentanol, water and dodecane, confirmed the theoretical predictions for the Landau–Peierls effect in the smectic-A phase (Safinya, Roux *et al.*, 1986). The problem of sample mosaic was resolved by using a three-dimensional powder. Although the conditions on the analysis are delicate, Safinya *et al.* demonstrated that for a perfect powder, for which the microcrystals are sufficiently large, the powder line shape does allow unambiguous determination of all of the parameters of the anisotropic line shape.

The only other X-ray study of a critical property on the smectic-A side of the transition has been a measurement of the temperature dependence of the integrated intensity of the peak. For three-dimensional crystals, the integrated intensity of a Bragg peak can be measured for samples with poor mosaic distributions, and because the differences between QLRO and true LRO are only manifest at long distances in real space, or at small wavevectors in reciprocal space, the same is true for the 'quasi-Bragg peak' of the smectic-A phase. Chan *et al.* measured the temperature dependence of the integrated intensity of the smectic-A peak across the nematic to smectic-A phase transition for a number of liquid crystals with varying exponents $\nu_\parallel$ and $\nu_\perp$ (Chan, Deutsch *et al.*, 1985). For the Landau–De Gennes free-energy density (equation 4.4.2.5), the theoretical prediction is that the critical part of the integrated intensity should vary as $|t|^x$, where $x = 1 - \alpha$ when the critical part of the heat capacity diverges according to the power law $|t|^{-\alpha}$. Six samples were measured with values of α varying from 0 to 0.5. Although for samples with $\alpha \approx 0.5$ the critical intensity did vary as $x \approx 0.5$, there were systematic deviations for smaller values of α, and for $\alpha \approx 0$ the measured values of x were in the range 0.7 to 0.76. The origin of this discrepancy is not at present understood.

Similar integrated intensity measurements in the vicinity of the first-order nematic to smectic-C transition cannot easily be made in smectic-C samples since the magnetic field aligns the molecular axis $\langle\mathbf{n}\rangle$, and when the layers form at some angle φ to $\langle\mathbf{n}\rangle$ the layer normals are distributed along the full 2π of azimuthal directions around $\langle\mathbf{n}\rangle$, as shown in Fig. 4.4.1.2(*d*). The X-ray scattering pattern for such a sample is a partial powder with a peak-intensity distribution that forms a ring of radius $|\mathbf{q}|\sin(\varphi)$. The opening of the single spot along the average molecular axis $\langle\mathbf{n}\rangle$ into a ring can be used to study either the nematic to smectic-C or the smectic-A to smectic-C transition (Martinez-Miranda *et al.*, 1986).

The statistical physics in the region of the phase diagram surrounding the triple point, where the nematic, smectic-A and smectic-C phases meet, has been the subject of considerable theoretical speculation (Chen & Lubensky, 1976; Chu &

McMillan, 1977; Benguigui, 1979; Huang & Lien, 1981; Grinstein & Toner, 1983). The best representation of the observed X-ray scattering structure near the nematic to smectic-A, the nematic to smectic-C and the nematic/smectic-A/smectic-C (NAC) multicritical point is obtained from the mean-field theory of Chen and Lubensky, the essence of which is expressed in terms of an energy density of the form

$$\Delta F(\psi) = \frac{A}{2}|\psi|^2 + \frac{D}{4}|\psi|^4 + \tfrac{1}{2}[E_\parallel(Q_\parallel^2 - Q_0^2)^2 + E_\perp Q_\perp^2 + E_{\perp\perp}Q_\perp^4 + E_{\perp\parallel}Q_\perp^2(Q_\parallel^2 - Q_0^2)]|\psi(\mathbf{Q})|^2, \quad (4.4.3.4)$$

where $\psi = \psi(\mathbf{Q})$ is the Fourier component of the electron density:

$$\psi(\mathbf{Q}) \equiv \frac{1}{(2\pi)^3}\int \mathrm{d}^3\mathbf{r}\exp[i(\mathbf{Q}\cdot\mathbf{r})]\rho(\mathbf{r}). \quad (4.4.3.5)$$

The quantities $E_\parallel$, $E_{\perp\perp}$, and $E_{\parallel\perp}$ are all positive definite; however, the sign of A and $E_\perp$ depends on temperature. For $A > 0$ and $E_\perp > 0$, the free energy, including the higher-order terms, is minimized by $\psi(\mathbf{Q}) = 0$ and the nematic is the stable phase. For $A < 0$ and $E_\perp > 0$, the minimum in the free energy occurs for a nonvanishing value for $\psi(\mathbf{Q})$ in the vicinity of $Q_\parallel \approx Q_0$, corresponding to the uniaxial smectic-A phase; however, for $E_\perp < 0$, the free-energy minimum occurs for a nonvanishing $\psi(\mathbf{Q})$ with a finite value of $Q_\perp$, corresponding to smectic-C order. The special point in the phase diagram where two terms in the free energy vanish simultaneously is known as a 'Lifshitz point' (Hornreich *et al.*, 1975). In the present problem, this occurs at the triple point where the nematic, smectic-A and smectic-C phases coexist. Although there have been other theoretical models for this transition, the best agreement between the observed and theoretical line shapes for the X-ray scattering cross sections is based on the Chen–Lubensky model. Most of the results from light-scattering experiments in the vicinity of the NAC triple point also agree with the main features predicted by the Chen–Lubensky model; however, there are some discrepancies that are not explained (Solomon & Litster, 1986).

The nematic to smectic-C transition in the vicinity of this point is particularly interesting in that, on approaching the nematic to smectic-C transition temperature from the nematic phase, the X-ray scattering line shapes first appear to be identical to the shapes usually observed on approaching the nematic to smectic-A phase transition; however, within approximately 0.1 K of the transition, they change to shapes that clearly indicate smectic-C-type fluctuations. Details of this crossover are among the strongest evidence supporting the Lifshitz idea behind the Chen–Lubensky model.

4.4.3.2. Modulated smectic-A and smectic-C phases

Previously, we mentioned that, although the reciprocal-lattice spacing $|\mathbf{q}|$ for many smectic-A phases corresponds to $2\pi/L$, where L is the molecular length, there are a number of others for which $|\mathbf{q}|$ is between π/L and $2\pi/L$ (Leadbetter, Frost, Gaughan, Gray & Mosley, 1979; Leadbetter *et al.*, 1977). This suggests the possibility of different types of smectic-A phases in which the bare molecular length is not the sole determining factor of the period d. In 1979, workers at Bordeaux optically observed some sort of phase transition between two phases that both appeared to be of the smectic-A type (Sigaud *et al.*, 1979). Subsequent X-ray studies indicated that in the nematic phase these materials simultaneously displayed critical fluctuations with two separate periods (Levelut *et al.*, 1981; Hardouin *et al.*, 1980, 1983; Ratna *et al.*, 1985, 1986; Chan, Pershan *et al.*, 1985, 1986; Safinya, Varady *et al.*, 1986; Fontes *et al.*, 1986) and confirmed phase transitions

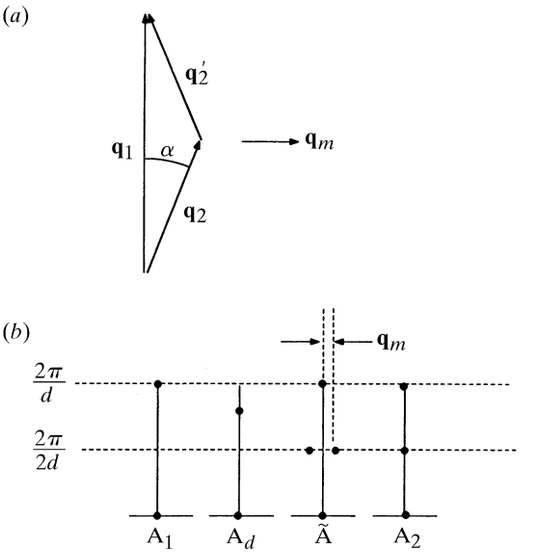


Fig. 4.4.3.1. (*a*) Schematic illustration of the necessary condition for coupling between order parameters when $|\mathbf{q}_2| < 2|\mathbf{q}_1|$; $|\mathbf{q}| = (|\mathbf{q}_2|^2 - |\mathbf{q}_1|^2)^{1/2} = |\mathbf{q}_1|\sin(\alpha)$. (*b*) Positions of the principal peaks for the indicated smectic-A phases.

between phases that have been designated smectic-A_1 with period $d \approx L$, smectic-A_2 with period $d \approx 2L$ and smectic-A_d with period $L < d < 2L$. Stimulated by the experimental results, Prost and co-workers generalized the De Gennes mean-field theory by writing

$$\rho(\mathbf{r}) = \langle\rho\rangle + \mathrm{Re}\{\Psi_1\exp(i\mathbf{q}_1\cdot\mathbf{r}) + \Psi_2\exp(i\mathbf{q}_2\cdot\mathbf{r})\},$$

where 1 and 2 refer to two different density waves (Prost, 1979; Prost & Barois, 1983; Barois *et al.*, 1985). In the special case that $\mathbf{q}_1 \approx 2\mathbf{q}_2$ the free energy represented by equation (4.4.2.3) must be generalized to include terms like

$$(\Psi_2^*)^2\Psi_1\exp[i(\mathbf{q}_1 - 2\mathbf{q}_2)\cdot\mathbf{r}] + \text{c.c.}$$

that couple the two order parameters. Suitable choices for the relative values of the phenomenological parameters of the free energy then result in minima that correspond to any one of these three smectic-A phases. Much more interesting, however, was the observation that even if $|\mathbf{q}_1| < 2|\mathbf{q}_2|$ the two order parameters could still be coupled together if $\mathbf{q}_1$ and $\mathbf{q}_2$ were not collinear, as illustrated in Fig. 4.4.3.1(*a*), such that $2\mathbf{q}_1\cdot\mathbf{q}_2 = |\mathbf{q}_1|^2$. Prost *et al.* predicted the existence of phases that are modulated in the direction perpendicular to the average layer normal with a period $4\pi/[|\mathbf{q}_2|\sin(\varphi)] = 2\pi/|\mathbf{q}_m|$. Such a modulated phase has been observed and is designated as the smectic-A (Hardouin *et al.*, 1981). Similar considerations apply to the smectic-C phases and the modulated phase is designated smectic-$\tilde{C}$; (Hardouin *et al.*, 1982; Huang *et al.*, 1984; Safinya, Varady *et al.*, 1986).

4.4.3.3. Surface effects

The effects of surfaces in inducing macroscopic alignment of mesomorphic phases have been important both for technological applications and for basic research (Sprokel, 1980; Gray & Goodby, 1984). Although there are a variety of experimental techniques that are sensitive to mesomorphic surface order (Beaglehole, 1982; Faetti & Palleschi, 1984; Faetti *et al.*, 1985; Gannon & Faber, 1978; Miyano, 1979; Mada & Kobayashi, 1981; Guyot-Sionnest *et al.*, 1986), it is only recently that X-ray scattering techniques have been applied to this problem. In one form or another, all of the techniques for obtaining surface specificity

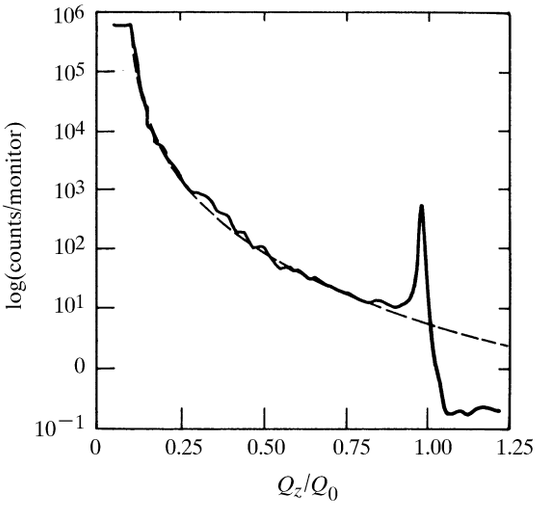


Fig. 4.4.3.2. Specular reflectivity of ∼8 keV X-rays from the air–liquid interface of the nematic liquid crystal 8OCB 0.05 K above the nematic to smectic-A transition temperature. The dashed line is the Fresnel reflection law as described in the text.

in an X-ray measurement make use of the fact that the average interaction between X-rays and materials can be treated by the introduction of a dielectric constant $\varepsilon \approx 1 - (4\pi\rho e^2/m\omega^2) = 1 - \rho r_e \lambda^2/\pi$, where ρ is the electron density, r_e is the classical radius of the electron, and ω and λ are the angular frequency and the wavelength of the X-ray. Since $\varepsilon < 1$, X-rays that are incident at a small angle to the surface θ_0 will be refracted in the material toward a smaller angle $\theta_T \approx (\theta_0^2 - \theta_c^2)^{1/2}$, where the 'critical angle' $\theta_c \approx (\rho r_e \lambda^2/\pi)^{1/2} \approx 0.003$ rad ($\approx 0.2^\circ$) for most liquid crystals (Warren, 1968). Although this is a small angle, it is at least two orders of magnitude larger than the practical angular resolution available in modern X-ray spectrometers (Als-Nielsen *et al.*, 1982; Pershan & Als-Nielsen, 1984; Pershan *et al.*, 1987). One can demonstrate that for many conditions the specular reflection $R(\theta_0)$ is given by

$$R(\theta_0) \approx R_F(\theta_0)|\rho^{-1} \int dz \exp(-iQz)\langle \partial\rho/\partial z\rangle|^2,$$

where $Q \equiv (4\pi/\lambda)\sin(\theta_0)$, $\langle \partial\rho/\partial z\rangle$ is the normal derivative of the electron density averaged over a region in the surface that is defined by the coherence area of the incident X-ray, and

$$R_F(\theta_0) \approx \left(\frac{\theta_0 - \sqrt{\theta_0^2 - \theta_c^2}}{\theta_0 + \sqrt{\theta_0^2 - \theta_c^2}}\right)^2$$

is the Fresnel reflection law that is calculated from classical optics for a flat interface between the vacuum and a material of dielectric constraint ε. Since the condition for specular reflection, that the incident and scattered angles are equal and in the same plane, requires that the scattering vector $\mathbf{Q} = \hat{z}(4\pi/\lambda)\sin(\theta_0)$ be parallel to the surface normal, it is quite practical to obtain, for flat surfaces, an unambiguous separation of the specular reflection signal from all other scattering events.

Fig. 4.4.3.2(*a*) illustrates the specular reflectivity from the free nematic–air interface for the liquid crystal 4′-octyloxybiphenyl-4-carbonitrile (8OCB) 0.050 K above the nematic to smectic-A phase-transition temperature (Pershan & Als-Nielsen, 1984). The dashed line is the Fresnel reflection $R_F(\theta_0)$ in units of $\sin(\theta_0)/\sin(\theta_c)$, where the peak at $\theta_c = 1.39^\circ$ corresponds to surface-induced smectic order in the nematic phase: *i.e.* the selection rule for specular reflection has been used to separate the specular reflection from the critical scattering from the bulk. Since the full width at half maximum is exactly equal to the reciprocal of the correlation length for critical fluctuations in the bulk, $2/\xi_\parallel$ at all temperatures from $T - T_{NA} \approx 0.006$ K up to values near to the nematic to isotropic transition, $T - T_{NA} \approx 3.0$ K, it is clear this is an example where the gravitationally induced long-range order in the surface position has induced mesomorphic order that has long-range correlations parallel to the surface. Along the surface normal, the correlations have only the same finite range as the bulk critical fluctuations. Studies on a number of other nematic (Gransbergen *et al.*, 1986; Ocko *et al.*, 1987) and isotropic surfaces (Ocko, Braslau *et al.*, 1986) indicate features that are specific to local structure of the surface.

4.4.4. Phases with in-plane order

Although the combination of optical microscopy and X-ray scattering studies on unoriented samples identified most of the mesomorphic phases, there remain a number of subtle features that were only discovered by spectra from well oriented samples (see the extensive references contained in Gray & Goodby, 1984). Nematic phases are sufficiently fluid that they are easily oriented by either external electric or magnetic fields, or surface boundary conditions, but similar alignment techniques are not generally successful for the more ordered phases because the combination of strains induced by thermal expansion and the enhanced elasticity that accompanies the order creates defects that do not easily anneal. Other defects that might have been formed during initial growth of the phase also become trapped and it is difficult to obtain well oriented samples by cooling from a higher-temperature aligned phase. Nevertheless, in some cases it has been possible to obtain crystalline-B samples with mosaic spreads of the order of a fraction of a degree by slowing cooling samples that were aligned in the nematic phase. In other cases, mesomorphic phases were obtained by heating and melting single crystals that were grown from solution (Benattar *et al.*, 1979; Leadbetter, Mazid & Malik, 1980).

Moncton & Pindak (1979) were the first to realize that X-ray scattering studies could be carried out on the freely suspended films that Friedel (1922) described in his classical treatise on liquid crystals. These samples, formed across a plane aperture (*i.e.* approximately 1 cm in diameter) in the same manner as soap bubbles, have mosaic spreads that are an order of magnitude smaller. The geometry is illustrated in Fig. 4.4.4.1(*a*). The substrate in which the aperture is cut can be glass (*e.g.* a microscope cover slip), steel or copper sheeting *etc.* A small amount of the material, usually in the high-temperature region of the smectic-A phase, is spread around the outside of an aperture that is maintained at the necessary temperature, and a wiper is used to drag some of the material across the aperture. If a stable film is successfully drawn, it is detected optically by its finite reflectivity. In particular, against a dark background and with the proper illumination it is quite easy to detect the thinnest free films.

In contrast to conventional soap films that are stabilized by electrostatic effects, smectic films are stabilized by their own layer structure. Films as thin as two molecular layers can be drawn and studied for weeks (Young *et al.*, 1978). Thicker films of the order of thousands of layers can also be made and, with some experience in depositing the raw material around the aperture and the speed of drawing, it is possible to draw films of almost any desired thickness (Moncton *et al.*, 1982). For films thinner than approximately 20 to 30 molecular layers (*i.e.* 600 to 1000 Å), the thickness is determined from the reflected intensity of a small helium–neon laser. Since the reflected intensities for films of 2, 3, 4, 5, ... layers are in the ratio of 4, 9, 16, 25, ..., the measurement can be calibrated by drawing and measuring a reasonable number of thin films. The most straightforward method for thick films is to measure the ellipticity of the polarization induced in laser light transmitted through the film at an

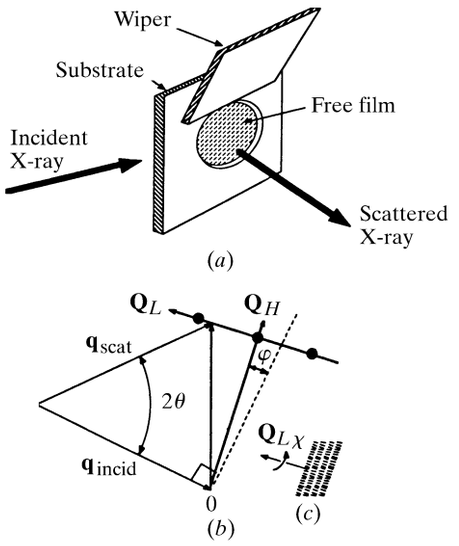


Fig. 4.4.4.1. (*a*) Schematic illustration of the geometry and (*b*) kinematics of X-ray scattering from a freely suspended smectic film. The insert (*c*) illustrates the orientation of the film in real space corresponding to the reciprocal-space kinematics in (*b*). If the angle $\varphi = \theta$, the film is oriented such that the scattering vector is parallel to the surface of the film, *i.e.* parallel to the smectic layers. A 'Q_L scan' is taken by simultaneous adjustment of φ and 2θ to keep $(4\pi/\lambda)\sin(\theta)\cos(\theta - \varphi) = (4\pi/\lambda)\sin(\theta_{100})$, where θ_{100} is the Bragg angle for the 100 reflection. The different in-plane Bragg reflections can be brought into the scattering plane by rotation of the film by the angle χ around the film normal.

oblique angle (Collett, 1983; Collett *et al.*, 1985); however, a subtler method that makes use of the colours of white light reflected from the films is also practical (Sirota, Pershan, Sorensen & Collett, 1987). In certain circumstances, the thickness can also be measured using the X-ray scattering intensity in combination with one of the other methods.

Fig. 4.4.4.1 illustrates the scattering geometry used with these films. Although recent unpublished work has demonstrated the possibility of a reflection geometry (Sorensen, 1987), all of the X-ray scattering studies to be described here were performed in transmission. Since the in-plane molecular spacings are typically between 4 and 5 Å, while the layer spacing is closer to 30 Å, it is difficult to study the 00*L* peaks in this geometry.

Fig. 4.4.4.2 illustrates the difference between X-ray scattering spectra taken on a bulk crystalline-B sample of *N*-[4-(*n*-butyloxy)benzylidene]-4-*n*-octylaniline (4O.8) that was oriented in an external magnetic field while in the nematic phase and then cooled through the smectic-A phase into the crystalline-B phase

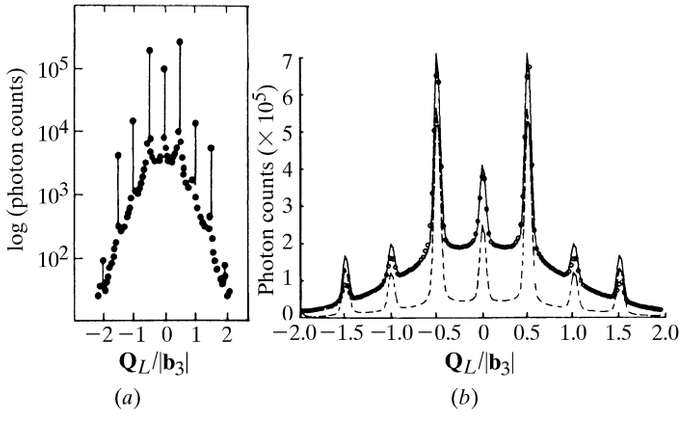


Fig. 4.4.4.2. Typical Q_L scans from the crystalline-B phases of (*a*) a free film of 7O.7, displayed on a logarithmic scale to illustrate the reduced level of the diffuse scattering relative to the Bragg reflection and (*b*) a bulk sample of 4O.8 oriented by a magnetic field.

(Aeppli *et al.*, 1981), and one taken on a thick freely suspended film of *N*-[4-(*n*-heptyloxy)benzylidene]-4-*n*-heptylaniline (7O.7) (Collett *et al.*, 1982, 1985). Note that the data for 7O.7 are plotted on a semi-logarithmic scale in order to display simultaneously both the Bragg peak and the thermal diffuse background. The scans are along the $\mathbf{Q}_L$ direction, at the appropriate value of $\mathbf{Q}_H$ to intersect the peaks associated with the intralayer periodicity. In both cases, the widths of the Bragg peaks are essentially determined by the sample mosaicity and as a result of the better alignment the ratio of the thermal diffuse background to the Bragg peak is nearly an order of magnitude smaller for the free film sample.

4.4.4.1. Hexatic phases in two dimensions

The hexatic phase of matter was first proposed independently by Halperin & Nelson (Halperin & Nelson 1978; Nelson & Halperin 1979) and Young (Young, 1979) on the basis of theoretical studies of the melting process in two dimensions. Following work by Kosterlitz & Thouless (1973), they observed that since the interaction energy between pairs of dislocations in two dimensions decreases logarithmically with their separation, the enthalpy and the entropy terms in the free energy have the same functional dependence on the density of dislocations. It follows that the free-energy difference between the crystalline and hexatic phase has the form $\Delta F = \Delta H - T\Delta \boldsymbol{S} \approx T_c \boldsymbol{S}(\rho) - T\boldsymbol{S}(\rho) = \boldsymbol{S}(\rho)(T_c - T)$, where $\boldsymbol{S}(\rho) \approx \rho \log(\rho)$ is the entropy as a function of the density of dislocations ρ and T_c is defined such that $T_c \boldsymbol{S}(\rho)$ is the enthalpy. Since the prefactor of the enthalpy term is independent of temperature while that of the entropy term is linear, there will be a critical temperature, T_c, at which the sign of the free energy changes from positive to negative. For temperatures greater than T_c, the entropy term will dominate and the system will be unstable against the spontaneous generation of dislocations. When this happens, the two-dimensional crystal, with positional QLRO, but true long-range order in the orientation of neighbouring atoms, can melt into a new phase in which the positional order is short range, but for which there is QLRO in the orientation of the six neighbours surrounding any atom. The reciprocal-space structures for the two-dimensional crystal and hexatic phases are illustrated in Figs. 4.4.4.3(*b*) and (*c*), respectively. That of the two-dimensional solid consists of a hexagonal lattice of sharp rods (*i.e.* algebraic line shapes in the plane of the crystal). For a finite size sample, the reciprocal-space structure of the two-dimensional hexatic phase is a hexagonal lattice of diffuse rods and there are theoretical predictions for the temperature dependence of the in-plane line shapes (Aeppli & Bruinsma, 1984). If the sample were of infinite size, the QLRO of the orientation would spread the six spots continuously around a circular ring, and the pattern would be indistinguishable from that of a well correlated liquid, *i.e.* Fig. 4.4.4.3(*a*). The extent of the patterns along the rod corresponds to the molecular form factor. Figs. 4.4.4.3(*a*), (*b*) and (*c*) are drawn on the assumption that the molecules are normal to the two-dimensional plane of the phase. If the molecules are tilted, the molecular form factor for long thin rod-like molecules will shift the intensity maxima as indicated in Figs. 4.4.4.3(*d*) and (*e*). The phase in which the molecules are normal to the two-dimensional plane is the two-dimensional *hexatic-B phase.* If the molecules tilt towards the position of their nearest neighbours (in real space), or in the direction that is between the lowest-order peaks in reciprocal space, the phase is the two-dimensional *smectic-I*, Fig. 4.4.4.3(*d*). The other tilted phase, for which the tilt direction is between the nearest neighbours in real space or in the direction of the lowest-order peaks in reciprocal space, is the *smectic-F*, Fig. 4.4.4.3(*e*).

Although theory (Halperin & Nelson, 1978; Nelson & Halperin, 1979; Young, 1979) predicts that the two-dimensional crystal can melt into a hexatic phase, it does not say that it must

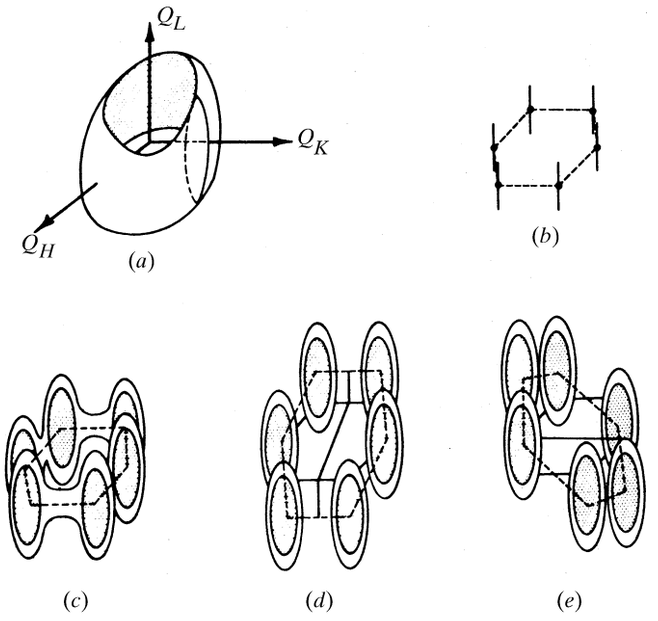


Fig. 4.4.4.3. Scattering intensities in reciprocal space from two-dimensional: (*a*) liquid; (*b*) crystal; (*c*) normal hexatic; and tilted hexatics in which the tilt is (*d*) towards the nearest neighbours as for the smectic-I or (*e*) between the nearest neighbours as for the smectic-F. The thin rods of scattering in (*b*) indicate the singular cusp for peaks with algebraic line shapes in the *HK* plane.

happen, and the crystal can melt directly into a two-dimensional liquid phase. Obviously, the hexatic phases will also melt into a two-dimensional liquid phase. Fig. 4.4.4.3(*a*) illustrates the reciprocal-space structure for the two-dimensional liquid in which the molecules are normal to the two-dimensional surface. Since the longitudinal (*i.e.* radial) width of the hexatic spot could be similar to the width that might be expected in a well correlated fluid, the direct X-ray proof of the transition from the hexatic-B to the normal liquid requires a hexatic sample in which the domains are sufficiently large that the sample is not a two-dimensional powder. On the other hand, the elastic constants must be sufficiently large that the QLRO does not smear the six spots into a circle. The radial line shape of the powder pattern of the hexatic-B phase can also be subtly different from that of the liquid and this is another possible way that X-ray scattering can detect melting of the hexatic-B phase (Aeppli & Bruinsma, 1984).

Changes that occur on the melting of the tilted hexatics, *i.e.* smectic-F and smectic-I, are usually easier to detect and this will be discussed in more detail below. On the other hand, there is a fundamental theoretical problem concerning the way of understanding the melting of the tilted hexatics. These phases actually have the same symmetry as the two-dimensional tilted fluid phase, *i.e.* the smectic-C. In two dimensions they all have QLRO in the tilt orientation, and since the simplest phenomenological argument says that there is a linear coupling between the tilt order and the near-neighbour positional order (Nelson & Halperin, 1980; Bruinsma & Nelson, 1981), it follows that the QLRO of the smectic-C tilt should induce QLRO in the near-neighbour positional order. Thus, by the usual arguments, if there is to be a phase transition between the smectic-C and one of the tilted hexatic phases, the transition must be a first-order transition (Landau & Lifshitz, 1958). This is analogous to the three-dimensional liquid-to-vapour transition which is first order up to a critical point, and beyond the critical point there is no real phase transition.

4.4.4.2. Hexatic phases in three dimensions

Based on both this theory and the various X-ray scattering patterns that had been reported in the literature (Gray & Goodby, 1984), Litster & Birgeneau (Birgeneau & Litster, 1978) suggested that some of the three-dimensional systems that were previously identified as mesomorphic were actually three-dimensional hexatic systems. They observed that it is not theoretically consistent to propose that the smectic phases are layers of two-dimensional crystals randomly displaced with respect to each other since, in thermal equilibrium, the interactions between layers of two-dimensional crystals must necessarily cause the layers to lock together to form a three-dimensional crystal.[1] On the other hand, if the layers were two-dimensional hexatics, then the interactions would have the effect of changing the QLRO of the hexagonal distribution of neighbours into the true long-range-order orientational distribution of the three-dimensional hexatic. In addition, interactions between layers in the three-dimensional hexatics can also result in interlayer correlations that would sharpen the width of the diffuse peaks in the reciprocal-space direction along the layer normal.

4.4.4.2.1. Hexatic-B

Although Leadbetter, Frost & Mazid (1979) had remarked on the different types of X-ray structures that were observed in materials identified as 'smectic-B', the first proof for the existence of the hexatic-B phase of matter was the experiment by Pindak *et al.* (1981) on thick freely suspended films of the liquid crystal *n*-hexyl 4′-pentyloxybiphenyl-4-carboxylate (65OBC). A second study on free films of the liquid crystal *n*-butyl 4′-*n*-hexyloxybiphenyl-4-carboxylate (46OBC) demonstrated that, as the hexatic-B melts into the smectic-A phase, the position and the in-plane width of the X-ray scattering peaks varied continuously. In particular, the in-plane correlation length evolved continuously from 160 Å, nearly 10 K below the hexatic to smectic-A transition, to only 17 Å, a few degrees above. Similar behaviour was also observed in a film only two layers thick (Davey *et al.*, 1984). Since the observed width of the peak along the layer normal corresponded to the molecular form factor, these systems have negligible interlayer correlations.

4.4.4.2.2. Smectic-F, smectic-I

In contrast to the hexatic-B phase, the principal reciprocal-space features of the smectic-F phase were clearly determined before the theoretical work that proposed the hexatic phase. Demus *et al.* (1971) identified a new phase in one material, and subsequent X-ray studies by Leadbetter and co-workers (Leadbetter, Mazid & Richardson, 1980; Leadbetter, Gaughan *et al.*, 1979; Gane & Leadbetter, 1981) and by Benattar and co-workers (Benattar *et al.*, 1978, 1980, 1983; Guillon *et al.*, 1986) showed it to have the reciprocal-space structure illustrated in Fig. 4.4.4.4(*b*). There are interlayer correlations in the three-dimensional smectic-F phases, and as a consequence the reciprocal-space structure has maxima along the diffuse rods. Benattar *et al.* (1979) obtained monodomain smectic-F samples of the liquid crystal *N*,*N*′-(1,4-phenylenedimethylene)bis(4-*n*-pentylaniline) by melting a single crystal that was previously precipitated from solution. One of the more surprising results of this work was the demonstration that the near-neighbour packing was very close to what would be expected from a model in which rigid closely packed rods were simply tilted away from the layer normal. In view of the facts that the molecules are clearly not cylindrical, and that the molecular tilt indicates that the macroscopic symmetry has been broken, it would have been reasonable to expect significant deviations from local hexagonal symmetry when the system is viewed along the molecular axis. The fact that this is not the case indicates that this phase has a considerable amount of rotational disorder around the long axis of the molecules.

[1] Prior to the paper by Birgeneau & Litster, it was commonly believed that some of the smectic phases consisted of uncorrelated stacks of two-dimensional crystals.

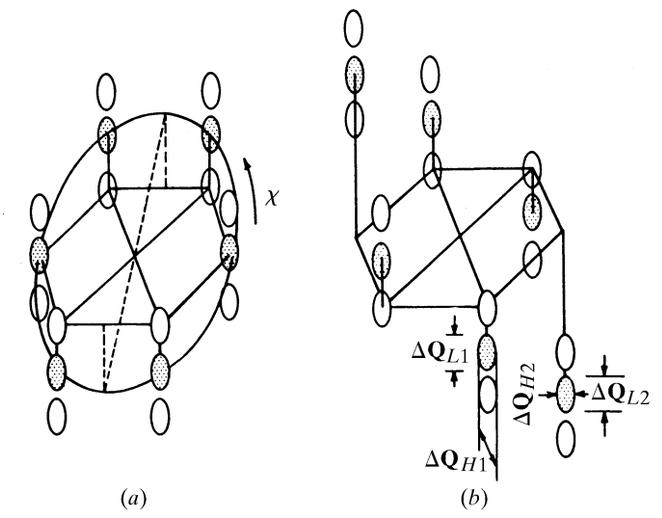


Fig. 4.4.4.4. Scattering intensities in reciprocal space from three-dimensional tilted hexatic phases: (*a*) the smectic-I and (*b*) the smectic-F. The variation of the intensity along the Q_L direction indicates interlayer correlations that are absent in Figs. 4.4.4.1(*d*) and (*e*). The peak widths $\Delta\mathbf{Q}_{L1,2}$ and $\Delta\mathbf{Q}_{H1,2}$ correspond to the four inequivalent widths in the smectic-F phase. Similar inequivalent widths exist for the smectic-I phase. The circle through the shaded points in (*a*) indicates the reciprocal-space scan that directly measures the hexatic order. A similar scan in the smectic-C phase would have intensity independent of χ.

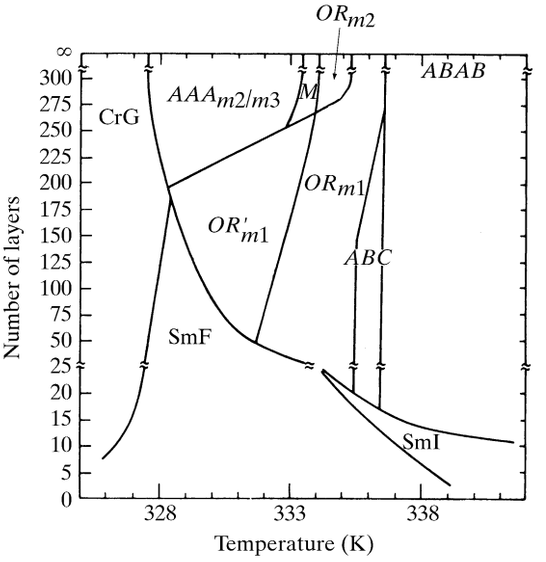


Fig. 4.4.4.5. The phase diagram for free films of 7O.7 as a function of thickness and temperature. The phases *ABAB*, *AAA*, OR_{m1}, OR_{m2}, OR'_{m1}, *M* and *ABAB* are all crystalline-B with varying interlayer stacking, or long-wavelength modulations; CrG, SmF and SmI are crystalline-G, smectic-F and smectic-I, respectively (Sirota *et al.*, 1985; Sirota, Pershan & Deutsch, 1987; Sirota, Pershan, Sorensen & Collett, 1987).

Other important features of the smectic-F phase are, firstly, that the local molecular packing is identical to that of the tilted crystalline-G phase (Benattar *et al.*, 1979; Sirota *et al.*, 1985; Guillon *et al.*, 1986). Secondly, there is considerable temperature dependence of the widths of the various diffuse peaks. Fig. 4.4.4.4(*b*) indicates the four inequivalent line widths that Sirota and co-workers measured in freely suspended films of the liquid crystal *N*-[4-(*n*-heptyloxy)benzylidene]-4-*n*-heptyl aniline (7O.7). Parenthetically, bulk samples of this material do not have a smectic-F phase; however, the smectic-F is observed in freely suspended films as thick as ∼200 layers. Fig. 4.4.4.5 illustrates the thickness–temperature phase diagram of 7O.7 between 325 and 342 K (Sirota *et al.*, 1985; Sirota, Pershan & Deutsch, 1987). Bulk samples and thick films have a first-order transition from the crystalline-B to the smectic-C at 342 K. Thinner films indicate a surface phase above 342 K that will be discussed below. Furthermore, although there is a strong temperature dependence of the widths of the diffuse scattering peaks, the widths are independent of film thickness. This demonstrates that, although the free film boundary conditions have stabilized the smectic-F phase, the properties of the phase are not affected by the boundaries. Finally, the fact that the widths $\Delta\mathbf{Q}_{L1}$ and $\Delta\mathbf{Q}_{L2}$ along the **L** direction and $\Delta\mathbf{Q}_{H1}$ and $\Delta\mathbf{Q}_{H2}$ along the in-plane directions are not equal indicates that the correlations are very anisotropic (Brock *et al.*, 1986; Sirota *et al.*, 1985). We will discuss one possible model for these properties after presenting other data on thick films of 7O.7. From the fact that the positions of the intensity maxima for the diffuse spots of the smectic-F phase of 7O.7 correspond exactly to the positions of the Bragg peaks in the crystalline-G phase, we learn that the local molecular packing must be identical in the two phases. The major difference between the crystalline-G and the tilted hexatic smectic-F phase is that, in the latter, defects destroy the long-range positional order of the former (Benattar *et al.*, 1979; Sirota *et al.*, 1985). Although this is consistent with the existing theoretical model that attributes hexatic order to a proliferation of unbounded dislocations, it is not obvious that the proliferation is attributable to the same Kosterlitz–Thouless mechanism that Halperin & Nelson and Young discussed for the transition from the two-dimensional crystal to the hexatic phase. We will say more on this point below.

The only identified difference between the two tilted hexatic phases, the smectic-F and the smectic-I, is the direction of the molecular tilt relative to the near-neighbour positions. For the smectic-I, the molecules tilt towards one of the near neighbours, while for the smectic-F they tilt between the neighbours (Gane & Leadbetter, 1983). There are a number of systems that have both smectic-I and smectic-F phases, and in all cases of which we are aware the smectic-I is the higher-temperature phase (Gray & Goodby, 1984; Sirota *et al.*, 1985; Sirota, Pershan, Sorensen & Collett, 1987).

Optical studies of freely suspended films of materials in the *n*O.*m* series indicated tilted surface phases at temperatures for which the bulk had uniaxial phases (Farber, 1985). As mentioned above, X-ray scattering studies of 7O.7 demonstrated that the smectic-F phase set in for a narrow temperature range in films as thick as 180 layers, and that the temperature range increases with decreasing layer number. For films of the order of 25 layers thick, the smectic-I phase is observed at approximately 334 K, and with decreasing thickness the temperature range for this phase also increases. Below approximately 10 to 15 layers, the smectic-I phase extends up to ∼342 K where bulk samples undergo a first-order transition from the crystalline-B to the smectic-C phase. Synchrotron X-ray scattering experiments show that, in thin films (five layers for example), the homogeneous smectic-I film undergoes a first-order transition to one in which the two surface layers are smectic-I and the three interior layers are smectic-C (Sirota *et al.*, 1985; Sirota, Pershan, Sorensen & Collett, 1987). The fact that two phases with the same symmetry can coexist in this manner tells us that in this material there is some important microscopic difference between them. This is reaffirmed by the fact that the phase transition from the surface smectic-I to the homogeneous smectic-C phase has been observed to be first order (Sorensen *et al.*, 1987).

In contrast to 7O.7, Birgeneau and co-workers found that in racemic 4-(2-methylbutyl)phenyl 4′-octyloxylbiphenyl-4-carboxylate (8OSI) (Brock *et al.*, 1986), the X-ray structure of the smectic-I phase evolves continuously into that of the smectic-C. By applying a magnetic field to a thick freely suspended sample, Brock *et al.* were able to obtain a large monodomain sample.

They measured the X-ray scattering intensity around the circle in the reciprocal-space plane shown in Fig. 4.4.4.4(*b*) that passes through the peaks. For higher temperatures, when the sample is in the smectic-C phase, the intensity is essentially constant around the circle; however, on cooling, it gradually condenses into six peaks, separated by 60°. The data were analysed by expressing the intensity as a Fourier series of the form

$$S(\chi) = I_0\left[\tfrac{1}{2} + \sum_{n=1}^{\infty} C_{6n}\cos 6n(90^\circ - \chi)\right] + I_{\rm B},$$

where I_0 fixes the absolute intensity and $I_{\rm B}$ fixes the background. The temperature variation of the coefficients scaled according to the relation $C_{6n} = C_6^{\sigma n}$ where the empirical relation $\sigma_n = 2.6(n-1)$ is in good agreement with a theoretical form predicted by Aharony *et al.* (1986). The only other system in which this type of measurement has been made was the smectic-C phase of 7O.7 (Collett, 1983). In that case, the intensity around the circle was constant, indicating the absence of any tilt-induced bond orientational order (Aharony *et al.*, 1986).

It would appear that the near-neighbour molecular packing of the smectic-I and the crystalline-J phases is the same, in just the same way as for the packing of the smectic-F and the crystalline-G phases. The four smectic-I widths analogous to those illustrated in Fig. 4.4.4.4(*a*) are, like that of the smectic-F, both anisotropic and temperature dependent (Sirota *et al.*, 1985; Sirota, Pershan, Sorensen & Collett, 1987; Brock *et al.*, 1986; Benattar *et al.*, 1979).

4.4.4.3. Crystalline phases with molecular rotation

4.4.4.3.1. Crystal-B

Recognition of the distinction between the hexatic-B and crystalline-B phases provided one of the more important keys to understanding the ordered mesomorphic phases. There are a number of distinct phases called crystalline-B that are all true three-dimensional crystals, with resolution-limited Bragg peaks (Moncton & Pindak, 1979; Aeppli *et al.*, 1981). The feature common to them all is that the average molecular orientation is normal to the layers, and within each layer the molecules are distributed on a triangular lattice. In view of the 'blade-like' shape of the molecule, the hexagonal site symmetry implies that the molecules must be rotating rapidly (Levelut & Lambert, 1971; Levelut, 1976; Richardson *et al.*, 1978). We have previously remarked that this apparent rotational motion characterizes all of the phases listed in Table 4.4.1.1 except for the crystalline-E, -H and -K. In the most common crystalline-B phase, adjacent layers have *ABAB*-type stacking (Leadbetter, Gaughan *et al.*, 1979; Leadbetter, Mazid & Kelly, 1979). High-resolution studies on well oriented samples show that in addition to the Bragg peaks the crystalline-B phases have rods of relatively intense diffuse scattering distributed along the 10*L* Bragg peaks (Moncton & Pindak, 1979; Aeppli *et al.*, 1981). The widths of these rods in the reciprocal-space direction, parallel to the layers, are very sharp, and without a high-resolution spectrometer their widths would appear to be resolution limited. In contrast, along the reciprocal-space direction normal to the layers, their structure corresponds to the molecular form factor.

If the intensity of the diffuse scattering can be represented as proportional to $\langle \mathbf{Q}\cdot\mathbf{u}\rangle^2$, where $\mathbf{u}$ describes the molecular displacement, the fact that there is no rod of diffuse scattering through the 00*L* peaks indicates that the rods through the 10*L* peaks originate from random disorder in 'sliding' displacements of adjacent layers. It is likely that these displacements are thermally excited phonon vibrations; however, we cannot rule out some sort of non-thermal static defect structure. In any event, assuming this diffuse scattering originates in a thermal vibration for which adjacent layers slide over one another with some amplitude $\langle \mathbf{u}^2\rangle^{1/2}$, and assuming strong coupling between this shearing motion and the molecular tilt, we can define an angle $\varphi = \tan^{-1}(\langle \mathbf{u}^2\rangle^{1/2}/d)$, where d is the layer thickness. The observed diffuse intensity corresponds to angles φ between 3 and 6° (Aeppli *et al.*, 1981).

Leadbetter and co-workers demonstrated that in the *n*O.*m* series various molecules undergo a series of restacking transitions and that crystalline-B phases exist with *ABC* and *AAA* stacking as well as the more common *ABAB* (Leadbetter, Mazid & Richardson 1980; Leadbetter, Mazid & Kelly, 1979). Subsequent high-resolution studies on thick freely suspended films revealed that the restacking transitions were actually subtler, and in 7O.7, for example, on cooling the hexagonal *ABAB* phase one observes an orthorhombic and then a monoclinic phase before the hexagonal *AAA* (Collett *et al.*, 1982, 1985). Furthermore, the first transition from the hexagonal *ABAB* to the monoclinic phase is accompanied by the appearance of a relatively long-wavelength modulation within the plane of the layers. The polarization of this modulation is along the layer normal, or orthogonal to the polarization of the displacements that gave rise to the rods of thermal diffuse scattering (Gane & Leadbetter, 1983).

It is also interesting to note that the *AAA* simple hexagonal structure does not seem to have been observed outside liquid-crystalline materials and, were it not for the fact that the crystalline-B hexagonal *AAA* is always accompanied by long-wavelength modulations, it would be the only case of which we are aware.

Figs. 4.4.4.6(*a*) and (*b*) illustrate the reciprocal-space positions of the Bragg peaks (dark dots) and modulation-induced side bands (open circles) for the unmodulated hexagonal *ABAB* and the modulated orthorhombic phase (Collett *et al.*, 1984). For convenience, we only display one 60° sector. Hirth *et al.* (1984) explained how both the reciprocal-space structure and the modulation of the orthorhombic phase could result from an ordered array of partial dislocations. They were not, however, able to provide a specific model for the microscopic driving force for the transition. Sirota, Pershan & Deutsch (1987) proposed a variation of the Hirth model in which the dislocations pair up to form a wall of dislocation dipoles such that within the wall the local molecular packing is essentially identical to the packing in the crystalline-G phase that appears at temperatures just below the crystalline-B phase. This model explains: (1) the macroscopic symmetry of the phase; (2) the period of the modulation; (3) the polarization of the modulation; and (4) the size of the observed deviations of the reciprocal-space structure from the hexagonal symmetry of the *ABAB* phase and suggests a microscopic driving mechanism that we will discuss below.

On further cooling, there is a first-order transition in which the one-dimensional modulation that appeared at the transition to orthorhombic symmetry is replaced by a two-dimensional modulation as shown in Fig. 4.4.4.6(*c*). On further cooling, there is another first-order transition in which the positions of the principal Bragg spots change from having orthorhombic to monoclinic symmetry as illustrated in Fig. 4.4.4.6(*d*). On further cooling, the Bragg peaks shift continually until there is one more first-order transition to a phase with hexagonal *AAA* positions as illustrated in Fig. 4.4.4.6(*f*). On further cooling, the *AAA* symmetry remains unchanged, and the modulation period is only slightly dependent on temperature, but the modulation amplitude increases dramatically. Eventually, as indicated in the phase diagram shown in Fig. 4.4.4.5, the system undergoes another first-order transition to the tilted crystalline-G phase. The patterns in Figs. 4.4.4.6(*e*) and (*g*) are observed by rapid quenching from the temperatures at which the patterns in Fig. 4.4.4.6(*b*) are observed.

Although there is not yet an established theoretical explanation for the origin of the 'restacking-modulation' effects, there are a number of experimental facts that we can summarize, and which indicate a probable direction for future research. Firstly, if one ignores the long-wavelength modulation, the hexagonal *ABAB* phase is the only phase in the diagram for 7O.7 for which

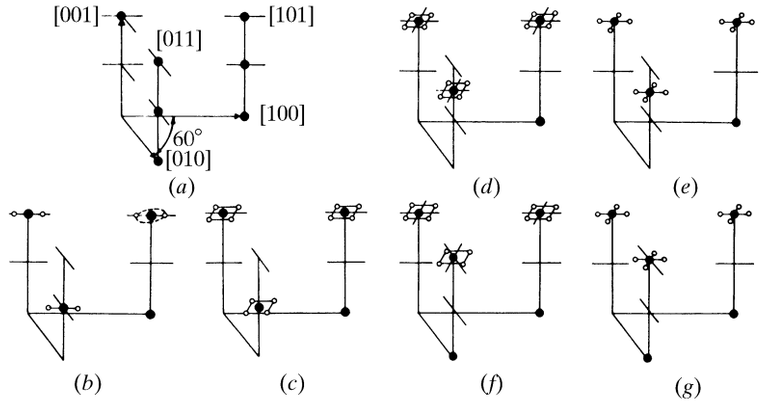


Fig. 4.4.4.6. Location of the Bragg peaks in one 60° section of reciprocal space for the three-dimensional crystalline-B phases observed in thick films of 7O.7. (*a*) The normal hexagonal crystalline-B phase with *ABAB* stacking. (*b*) The one-dimensional modulated phase with orthorhombic symmetry. The closed circles are the principal Bragg peaks and the open circles indicate side bands associated with the long-wavelength modulation. (*c*) The two-dimensional modulated phase with orthorhombic symmetry. Only the lowest-order side bands are shown. They are situated on the corners of squares surrounding the Bragg peak. The squares are oriented as shown and the amplitude of the square diagonal is equal to the distance between the two side bands illustrated in (*b*). (*d*) The two-dimensional modulated phase with monoclinic symmetry. Note that the *L* position of one of the peaks has shifted relative to (*c*). (*e*) A two-dimensional modulated phase with orthorhombic symmetry that is only observed on heating the quenched phase illustrated in (*g*). (*f*) The two-dimensional modulated phase with hexagonal symmetry and *AAA* layer stacking. (*g*) A two-dimensional hexagonal phase with *AAA* layer stacking that is only observed on rapid cooling from the phase shown in (*c*).

there are two molecules per unit cell. There must be some basic molecular effect that determines this particular coupling between every other layer. In addition, it is particularly interesting that it only manifests itself for a small temperature range and then vanishes as the sample is cooled. Secondly, any explanation for the driving force of the restacking transition must also explain the modulations that accompany it. In particular, unless one cools rapidly, the same modulation structures with the same amplitudes always appear at the same temperature, regardless of the sample history, *i.e.* whether heating or cooling. No significant hysteresis is observed and Sirota argued that the structures are in thermal equilibrium.

There are a number of physical systems for which the development of long-wavelength modulations is understood, and in each case they are the result of two or more competing interaction energies that cannot be simultaneously minimized (Blinc & Levanyuk, 1986; Safinya, Varady *et al.*, 1986; Lubensky *et al.*, 1988; Winkor & Clarke, 1986; Moncton *et al.*, 1981; Fleming *et al.*, 1980; Villain, 1980; Frank & van der Merwe, 1949; Bak *et al.*, 1979; Pokrovsky & Talapov, 1979). The easiest to visualize is epitaxic growth of one crystalline phase on the surface of another when the two lattice vectors are slightly incommensurate. The first atomic row of adsorbate molecules can be positioned to minimize the attractive interactions with the substrate. This is slightly more difficult for the second row, since the distance that minimizes the interaction energy between the first and second rows of adsorbate molecules is not necessarily the same as the distance that would minimize the interaction energy between the first row and the substrate. As more and more rows are added, the energy price of this incommensurability builds up, and one possible configuration that minimizes the global energy is a modulated structure.

In all known cases, the very existence of modulated structures implies that there must be competing interactions, and the only real question about the modulated structures in the crystalline-B phases is the identification of the competing interactions. It appears that one of the more likely possibilities is the difficulty in packing the 7O.7 molecules within a triangular lattice while simultaneously optimizing the area per molecule of the alkane tails and the conjugated rings in the core (Carlson & Sethna, 1987; Sadoc & Charvolin, 1986). Typically, the mean cross-sectional area for a straight alkane in the all-*trans* configuration is between 18 and 19 Å^2, while the mean area per molecule in the crystalline-B phase is closer to 24 Å^2. While these two could be reconciled by assuming that the alkanes are tilted with respect to the conjugated core, there is no reason why the angle that reconciles the two should also be the same angle that minimizes the internal energy of the molecule. Even if it were the correct angle at some temperature by accident, the average area per chain is certainly temperature dependent. Even without attempting to include the rotational dynamics that are necessary to understanding the axial site symmetry, it is obvious that there can be a conflict in the packing requirements of the two different parts of the molecule.

A possible explanation of these various structures might be as follows: at high temperatures, both the alkane chain, as well as the other degrees of freedom, have considerable thermal motions that make it possible for the conflicting packing requirements to be simultaneously reconciled by one or another compromise. On the other hand, with decreasing temperature, some of the thermal motions become frozen out, and the energy cost of the reconciliation that was possible at higher temperatures becomes too great. At this point, the system must find another solution, and the various modulated phases represent the different compromises. Finally, all of the compromises involving inhomogeneities, like the modulations or grain boundaries, become impossible and the system transforms into a homogeneous crystalline-G phase.

If this type of argument could be made more specific, it would also provide a possible explanation for the molecular origin of the three-dimensional hexatic phases. The original suggestion for the existence of hexatic phases in two dimensions was based on the fact that the interaction energy between dislocations in two dimensions was logarithmic, such that the entropy and the enthalpy had the same functional dependence on the density of dislocations. This gave rise to the observation that above a certain temperature two-dimensional crystals would be unstable against thermally generated dislocations. Although Litster and Birgeneau's suggestion that some of the observed smectic phases might be stacks of two-dimensional hexatics is certainly correct, it is not necessary that the observed three-dimensional hexatics originate from entropy-driven thermally excited dislocations. For example, the temperature–layer-number phase diagram for 7O.7 that is shown in Fig. 4.4.4.5 has the interesting property that the temperature region over which the tilted hexatic phases exist in thin films is almost the same as the temperature region for which the modulated phases exist in thick films and in bulk samples.

From the fact that molecules in the *n*O.*m* series that only differ by one or two $-CH_2-$ groups have different sequences of mesomorphic phases, we learn that within any one molecule the

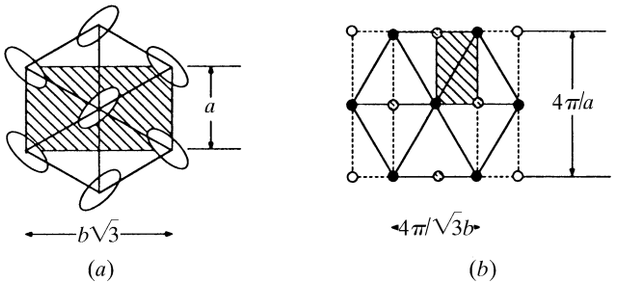


Fig. 4.4.4.7. (*a*) The 'herringbone' stacking suggested for the crystalline-E phase in which molecular rotation is partially restricted. The primitive rectangular unit cell containing two molecules is illustrated by the shaded region. The lattice has rectangular symmetry and $a \neq b$. (*b*) The position of the Bragg peaks in the plane in reciprocal space that is parallel to the layers. The dark circles indicate the principal Bragg peaks that would be the only ones present if all molecules were equivalent. The open circles indicate additional peaks that are observed for the model illustrated in (*a*). The cross-hatched circles indicate peaks that are missing because of the glide plane in (*a*).

difference in chemical potentials between the different mesomorphic phases must be very small (Leadbetter, Mazid & Kelly, 1979; Doucet & Levelut, 1977; Leadbetter, Frost & Mazid, 1979; Leadbetter, Mazid & Richardson, 1980; Smith *et al.*, 1973; Smith & Garland, 1973). For example, although in 7O.7 the smectic-F phase is only observed in finite-thickness films, both 5O.6 and 9O.4 have smectic-F phase in bulk. Thus, in bulk 7O.7 the chemical potential for the smectic-F phase must be only slightly larger than that of the modulated crystalline-B phases, and the effect of the surfaces must be sufficient to reverse the order in samples of finite thickness.

As far as the appearance of the smectic-F phase in 7O.7 is concerned, it is well known that the interaction energy between dislocation pairs is very different near a free surface from that in the bulk (Pershan, 1974; Pershan & Prost, 1975). The origin of this is that the elastic properties of the surface will usually cause the stress field of a dislocation near to the surface either to vanish or to be considerably smaller than it would in the bulk. Since the interaction energy between dislocations depends on this stress field, the surface significantly modifies the dislocation–dislocation interaction. This is a long-range effect, and it would not be surprising if the interactions that stabilized the dislocation arrays to produce the long-wavelength modulations in the thick samples were sufficiently weaker in the samples of finite thickness that the dislocation arrays are disordered. Alternatively, there is evidence that specific surface interactions favour a finite molecular tilt at temperatures where the bulk phases are uniaxial (Farber, 1985). Incommensurability between the period of the tilted surface molecules and the crystalline-B phases below the surface would increase the density of dislocations, and this would also modify the dislocation–dislocation interactions in the bulk.

Sirota *et al.* (1985) and Sirota, Pershan, Sorensen & Collett (1987) demonstrated that, while the correlation lengths of the smectic-F phase have a significant temperature dependence, the lengths are independent of film thickness, and this supports the argument that although the effects of the surface are important in stabilizing the smectic-F phase in 7O.7, once the phase is established it is essentially no different from the smectic-F phases observed in bulk samples of other materials. Brock *et al.* (1986) observed anisotropies in the correlation lengths of thick samples of 8OSI that are similar to those observed by Sirota.

These observations motivate the hypothesis that the dislocation densities in the smectic-F phases are determined by the same incommensurability that gives rise to the modulated crystalline-B structures. Although all of the experimental evidence supporting this hypothesis was obtained from the smectic-F tilted hexatic phase, there is no reason why this speculation could not apply to both the tilted smectic-I and the untilted hexatic-B phase.

4.4.4.3.2. Crystal-G, crystal-J

The crystalline-G and crystalline-J phases are the ordered versions of the smectic-F and smectic-I phases, respectively. The positions of the principal peaks illustrated in Fig. 4.4.4.4 for the smectic-F(I) are identical to the positions in the smectic-G(J) phase if small thermal shifts are discounted. In both the hexatic and the crystalline phases, the molecules are tilted with respect to the layer normals by approximately 25 to 30° with nearly hexagonal packing around the tilted axis (Doucet & Levelut, 1977; Levelut *et al.*, 1974; Levelut, 1976; Leadbetter, Mazid & Kelly, 1979; Sirota, Pershan, Sorensen & Collett, 1987). The interlayer molecular packing appears to be end to end, in an *AAA* type of stacking (Benattar *et al.*, 1983, 1981; Levelut, 1976; Gane *et al.*, 1983). There is only one molecule per unit cell and there is no evidence for the long-wavelength modulations that are so prevalent in the crystalline-B phase that is the next higher temperature phase above the crystalline-G in 7O.7.

4.4.4.4. Crystalline phases with herringbone packing

4.4.4.4.1. Crystal-E

Fig. 4.4.4.7 illustrates the intralayer molecular packing proposed for the crystalline-E phase (Levelut, 1976; Doucet, 1979; Levelut *et al.*, 1974; Doucet *et al.*, 1975; Leadbetter *et al.*, 1976; Richardson *et al.*, 1978; Leadbetter, Frost, Gaughan & Mazid, 1979; Leadbetter, Frost & Mazid, 1979). The molecules are, on average, normal to the layers; however, from the optical birefringence it is apparent that the site symmetry is not uniaxial. X-ray diffraction studies on single crystals by Doucet and co-workers demonstrated that the biaxiality was not attributable to molecular tilt and subsequent work by a number of others resulted in the arrangement shown in Fig. 4.4.4.7(*a*). The most important distinguishing reciprocal-space feature associated with the intralayer 'herringbone' packing is the appearance of Bragg peaks at $\sin(\theta)$ equal to $\sqrt{7}/2$ times the value for the lowest-order in-plane Bragg peak for the triangular lattice (Pindak *et al.*, 1981). These are illustrated by the open circles in Fig. 4.4.4.7(*b*). The shaded circles correspond to peaks that are missing because of the glide plane that relates the two molecules in the rectangular cell.

Leadbetter, Mazid & Malik (1980) carried out detailed studies on both the crystalline-E phase of isobutyl 4-(4-phenylbenzylideneamino)cinnamate (IBPBAC) and the crystalline phase immediately below the crystalline-E phase. Partially ordered samples of the crystalline-E phase were obtained by melting the lower-temperature crystalline phase. Although the data for the crystalline-E phase left some ambiguity, they argued that the phase they were studying might well have had molecular tilts of the order of 5 or 6°. This is an important distinction, since the crystalline-H and crystalline-J phases are essentially tilted versions of the crystalline-E. Thus, one important symmetry difference that might distinguish the crystalline-E from the others is the presence of a mirror plane parallel to the layers. In view of the low symmetry of the individual molecules, the existence of such a mirror plane would imply residual molecular motions. In fact, using neutron diffraction Leadbetter *et al.* (1976) demonstrated for a different liquid crystal that, even though the site symmetry is not axially symmetric, there is considerable residual rotational motion in the crystalline-E phase about the long axis of the molecules. Since the in-plane spacing is too small for neighbouring molecules to be rotating independently of each other, they proposed what might be interpreted as large partially hindered rotations.

4.4.4.4.2. Crystal-H, crystal-K

The crystalline-H and crystalline-K phases are tilted versions of the crystalline-E. The crystalline-H is tilted in the direction between the near neighbours, with the convenient mnemonic that

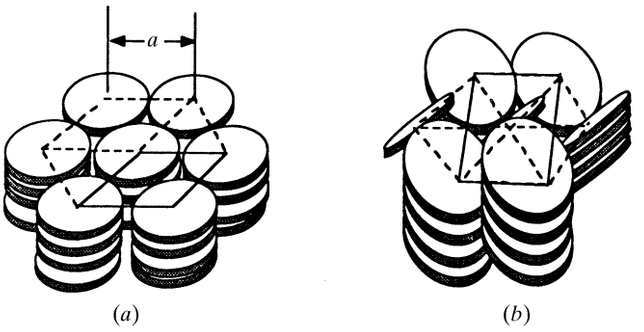


Fig. 4.4.5.1. Schematic illustration of the molecular stacking for the discotic (*a*) D_2 and (*b*) D_1 phases. In neither of these two phases is there any indication of long-range positional order along the columns. The hexagonal symmetry of the D_1 phase is broken by 'herringbone-like' correlations in the molecular tilt from column to column.

on cooling the sequence of phases with the same relative orientation of tilt to near-neighbour position is F → G → H. Similarly, the tilt direction for the crystalline-K phase is similar to that of the smectic-I and crystalline-J so that the expected phase sequence on cooling might be I → J → K. In fact, both of these sequences are only intended to indicate the progression in lower symmetry; the actual transitions vary from material to material.

4.4.5. Discotic phases

In contrast to the long thin rod-like molecules that formed most of the other phases discussed in this chapter, the discotic phases are formed by molecules that are more disc-like [see Fig. 4.4.1.3(*f*), for example]. There was evidence that mesomorphic phases were formed from disc-like molecules as far back as 1960 (Brooks & Taylor, 1968); however, the first identification of a discotic phase was by Chandrasekhar *et al.* (1977) with benzenehexyl hexa-*n*-alkanoate compounds. Disc-like molecules can form either a fluid nematic phase in which the disc normals are aligned, without any particular long-range order at the molecular centre of mass, or more-ordered 'columnar' (Helfrich, 1979) or 'discotic' (Billard *et al.*, 1981) phases in which the molecular positions are correlated such that the discs stack on top of one another to form columns. Some of the literature designates this nematic phase as N_D to distinguish it from the phase formed by 'rod-like' molecules (Destrade *et al.*, 1983). In the same way that the appearance of layers characterizes order in smectic phases, the order for the discotic phases is characterized by the appearance of columns. Chandrasekhar (1982, 1983) and Destrade *et al.* (1983) have reviewed this area and have summarized the several notations for various phases that appear in the literature. Levelut (1983) has also written a review and presented a table listing the space groups for columnar phases formed by 18 different molecules. Unfortunately, it is not absolutely clear which of these are mesomorphic phases and which are crystals with true long-range positional order.

Fig. 4.4.5.1 illustrates the molecular packing in two of the well identified discotic phases that are designated as D_1 and D_2 (Chandrasekhar, 1982). The phase D_2 consists of a hexagonal array of columns for which there is no intracolumnar order. The system is uniaxial and, as originally proposed, the molecular normals were supposed to be along the column axis. However, recent X-ray scattering studies on oriented free-standing fibres of the D_2 phase of triphenylene hexa-*n*-dodecanoate indicate that the molecules are tilted with respect to the layer normal (Safinya *et al.*, 1985, 1984). The D_1 phase is definitely a tilted phase, and consequently the columns are packed in a rectangular cell. According to Safinya *et al.*, the D_1 to D_2 transition corresponds to an order–disorder transition in which the molecular tilt orientation is ordered about the column axis in the D_1 phase and disordered in the D_2 phase. The reciprocal-space structure of the D_1 phase is similar to that of the crystalline-E phase shown in Fig. 4.4.4.7(*b*).

Other discotic phases that have been proposed would have the molecules arranged periodically along the column, but disordered between columns. This does not seem physically realistic since it is well known that thermal fluctuations rule out the possibility of a one-dimensional periodic structure even more strongly than for the two-dimensional lattice that was discussed above (Landau, 1965; Peierls, 1934). On the other hand, in the absence of either more high-resolution studies on oriented fibres or further theoretical studies, we prefer not to speculate on the variety of possible true discotic or discotic-like crystalline phases that might exist. This is a subject for future research.

4.4.6. Other phases

We have deliberately chosen not to discuss the properties of the cholesteric phase in this chapter because the length scales that characterize the long-range order are of the order of micrometres and are more easily studied by optical scattering than by X-rays (De Gennes, 1974; De Vries, 1951). Nematic phases formed from chiral molecules develop long-range order in which the orientation of the director $\langle \mathbf{n} \rangle$ varies in a plane-wave-like manner that can be described as $\mathbf{x}\cos(2\pi z/\lambda) + \mathbf{y}\sin(2\pi z/\lambda)$, where $\mathbf{x}$ and $\mathbf{y}$ are unit vectors and $\lambda/2$ is the cholesteric 'pitch' that can be anywhere from 0.1 to 10 μm depending on the particular molecule. Even more interesting is that for many cholesteric systems there is a small temperature range, of the order of 1 K, between the cholesteric and isotropic phases for which there is a phase known as the 'blue phase' (Coates & Gray, 1975; Stegemeyer & Bergmann, 1981; Meiboom *et al.*, 1981; Bensimon *et al.*, 1983; Hornreich & Shtrikman, 1983; Crooker, 1983). In fact, there is more than one 'blue phase' but they all have the property that the cholesteric twist forms a three-dimensional lattice twisted network rather than the plane-wave-like twist of the cholesteric phase. Three-dimensional Bragg scattering from blue phases using laser light indicates cubic lattices; however, since the optical cholesteric interactions are much stronger than the usual interactions between X-rays and atoms, interpretation of the results is subtler.

Gray and Goodby discuss a 'smectic-D' phase that is otherwise omitted from this chapter (Gray & Goodby, 1984). Gray and coworkers first observed this phase in the homologous series of 4′-*n*-alkoxy-3′-nitrobiphenyl-4-carboxylic acids (Gray *et al.*, 1957). In the hexadecyloxy compound, this phase exists for a region of about 26 K between the smectic-C and smectic-A phases: smectic-C (444.2 K) smectic-D (470.4 K) smectic-A. It is optically isotropic and X-ray studies by Diele *et al.* (1972) and by Tardieu & Billard (1976) indicate a number of similarities to the 'cubic–isotropic' phase observed in lyotropic systems (Luzzati & Riess-Husson, 1966; Tardieu & Luzzati, 1970). More recently, Etherington *et al.* (1986) studied the 'smectic-D' phase of 3′-cyano-4′-*n*-octadecyloxybiphenyl-4-carboxylic acid. Since this material appears to be more stable than some of the others that were previously studied, they were able to perform sufficient measurements to determine that the space group is cubic $P23$ or $Pm3$ with a lattice parameter of 86 Å. Etherington *et al.* suggested that the 'smectic-D' phase that they studied is a true three-dimensional cubic crystal of micelles and noted that the designation of 'smectic-D' is not accurate. Guillon & Skoulios (1987) have proposed a molecular model for this and related phases.

Fontell (1974) has reviewed the literature on the X-ray diffraction studies of lyotropic mesomorphic systems and the reader is referred there for more extensive information on those cubic systems. The mesomorphic structures of lyotropic systems are much richer than those of the thermotropic and, in addition

to all structures mentioned here, there are lyotropic systems in which the smectic-A lamellae seem to break up into cylindrical rods which seem to have the same macroscopic symmetry as some of the discotic phases. On the other hand, it is also much more difficult to prepare a review for the lyotropic systems in the same type of detail as for the thermotropic. The extra complexity associated with the need to control water concentration as well as temperature has made both theoretical and experimental progress more difficult, and, since there has not been very much experimental work on well oriented samples, detailed knowledge of many of these phases is also limited. Aside from the simpler lamellae systems, which seem to have the same symmetry as the thermotropic smectic-A phase, it is not at all clear which of the other phases are three-dimensional crystals and which are true mesomorphic structures. For example, dipalmitoylphosphatidylcholine has an L_β phase that appears for temperatures and (or) water content that is lower than that of the smectic-A L_α phase (Shipley *et al.*, 1974; Small, 1967; Chapman *et al.*, 1967). The diffraction pattern for this phase contains sharp large-angle reflections that may well correspond to a phase that is like one of the crystalline phases listed in Tables 4.4.1.1 and 4.4.1.2, and Fig. 4.4.1.1. On the other hand, this phase could also be hexatic and we do not have sufficient information to decide. The interested reader is referred to the referenced articles for further detailed information.

4.4.6.1. Phases with intermediate molecular tilt: smectic-L, crystalline-M,N

Following the completion of this manuscript, Smith and coworkers (Smith *et al.*, 1988; Sirota *et al.*, 1988) published an X-ray scattering study of the structure of a freely suspended multilayer film of hydrated phosphatidylcholine in which the phase that had been designated $L_{B'}$ in the literature on lipid phases (Janiak *et al.*, 1979; Luzzati, 1968; Tardieu *et al.*, 1973) was shown to consist of three separate two-dimensional phases in which the positional order in adjacent layers is uncoupled. The three phases are distinguished by the direction of the alkane-chain tilt relative to the nearest neighbours, and in one of these phases the orientation varies continuously with increasing hydration. At the lowest hydration, they observe a phase in which the tilt is towards the second-nearest neighbour; in analogy to the smectic-F phase, they designate this phase $L_{\beta F}$. On increasing hydration, they observe a phase in which the tilt direction is intermediate between the nearest- and next-nearest-neighbour directions, and which varies continuously with hydration. This is a new phase that was not previously known and they designate it $L_{\beta L}$. On further hydration, they observe a phase in which the molecular tilt is towards a nearest neighbour and this is designated $L_{\beta I}$. At maximum hydration, they observe the phase with long-wavelength modulation that was previously designated P_β (Janiak *et al.*, 1979). Selinger & Nelson (1988) have subsequently developed a theory for the phase transitions between phases with varying tilt orientation and have rationalized the existence of phases with intermediate tilt. To be complete, both Fig. 4.4.1.1 and Table 4.4.1.1 should be amended to include this type of hexatic order which is now referred to as the smectic-L. Extension of the previous logic suggests that the crystalline phases with intermediate tilt should be designated M and N, where N has 'herringbone' type of intermolecular order.

4.4.6.2. Nematic to smectic-A phase transition

At the time this manuscript was prepared, there was a fundamental discrepancy between theoretical predictions for the details of the critical properties of the second-order nematic to smectic-A phase transition. This has been resolved. Bouwman & de Jeu (1992) reported an X-ray scattering study of the critical properties of octyloxyphenylcyanobenzyloxybenzoate in which the data were in good agreement with predictions of the three-dimensional *xy* model (Lubensky, 1983; Le Guillou & Zinn-Justin, 1985). The differences between this experiment and others that were discussed previously, and which did not agree with theory, are firstly that this material is much further from the tricritical point that appears to be ubiquitous for most liquid-crystalline materials and, secondly, that they used the Landau–De Gennes theory to argue that the critical-temperature dependence for the $\mathbf{Q}_\perp^4$ term in the differential cross section given in equation (4.4.2.7) is not that of the $c\xi_\perp^4$ term but rather should vary as $[(T - T_{NA})/T]^{-\gamma/4}$, where γ is the exponent that describes the critical-temperature dependence of the smectic order parameter $|\Psi|^2 \simeq [(T - T_{NA})/T]^{-\gamma}$. The experimental results are in good agreement with the Monte Carlo simulation of the N–S_A transition that was reported by Dasgupta (1985, 1987).

References

Aeppli, G. & Bruinsma, R. (1984). *Hexatic order and liquid crystal density fluctuations. Phys. Rev. Lett.* **53**, 2133–2136.

Aeppli, G., Litster, J. D., Birgeneau, R. J. & Pershan, P. S. (1981). *High resolution X-ray study of the smectic A–smectic B phase transition and the smectic B phase in butyloxybenzylidene octylaniline. Mol. Cryst. Liq. Cryst.* **67**, 205–214.

Aharony, A., Birgeneau, R. J., Brock, J. D. & Litster, J. D. (1986). *Multicriticality in hexatic liquid crystals. Phys. Rev. Lett.* **57**, 1012–1015.

Alben, R. (1973). *Phase transitions in a fluid of biaxial particles. Phys. Rev. Lett.* **30**, 778–781.

Als-Nielsen, J., Christensen, F. & Pershan, P. S. (1982). *Smectic-A order at the surface of a nematic liquid crystal: synchrotron X-ray diffraction. Phys. Rev. Lett.* **48**, 1107–1110.

Als-Nielsen, J., Litster, J. D., Birgeneau, R. J., Kaplan, M., Safinya, C. R., Lindegaard, A. & Mathiesen, S. (1980). *Observation of algebraic decay of positional order in a smectic liquid crystal. Phys. Rev. B*, **22**, 312–320.

Bak, P., Mukamel, D., Villain, J. & Wentowska, K. (1979). *Commensurate–incommensurate transitions in rare-gas monolayers adsorbed on graphite and in layered charge-density-wave systems. Phys. Rev. B*, **19**, 1610–1613.

Barois, P., Prost, J. & Lubensky, T. C. (1985). *New critical points in frustrated smectics. J. Phys. (Paris)*, **46**, 391–399.

Beaglehole, D. (1982). *Pretransition order on the surface of a nematic liquid crystal. Mol. Cryst. Liq. Cryst.* **89**, 319–325.

Benattar, J. J., Doucet, J., Lambert, M. & Levelut, A. M. (1979). *Nature of the smectic F phase. Phys. Rev. A*, **20**, 2505–2509.

Benattar, J. J., Levelut, A. M. & Strzelecki, L. (1978). *Etude de l'influence de la longueur moléculaire les charactéristiques des phases smectique ordonnées. J. Phys. (Paris)*, **39**, 1233–1240.

Benattar, J. J., Moussa, F. & Lambert, M. (1980). *Two-dimensional order in the smectic F phase. J. Phys. (Paris)*, **40**, 1371–1374.

Benattar, J. J., Moussa, F. & Lambert, M. (1983). *Two dimensional ordering in liquid crystals: the SmF and SmI phases. J. Chim. Phys.* **80**, 99–107.

Benattar, J. J., Moussa, F., Lambert, M. & Germian, C. (1981). *Two kinds of two-dimensional order: the SmF and SmI phases. J. Phys. (Paris) Lett.* **42**, L67–L70.

Benguigui, L. (1979). *A Landau theory of the NAC point. J. Phys. (Paris) Colloq.* **40**, C3-419–C3-421.

Bensimon, D., Domany, E. & Shtrikman, S. (1983). *Optical activity of cholesteric liquid crystals in the pretransitional regime and in the blue phase. Phys. Rev. A*, **28**, 427–433.

Billard, J., Dubois, J. C., Vaucher, C. & Levelut, A. M. (1981). *Structures of the two discophases of rufigallol hexa-n-octanoate. Mol. Cryst. Liq. Cryst.* **66**, 115–122.

Birgeneau, R. J., Garland, C. W., Kasting, G. B. & Ocko, B. M. (1981). *Critical behavior near the nematic–smectic-A transition in butyloxybenzilidene octylaniline (4O.8). Phys. Rev. A*, **24**, 2624–2634.

Birgeneau, R. J. & Litster, J. D. (1978). *Bond orientational order model for smectic B liquid crystals. J. Phys. (Paris) Lett.* **39**, L399–L402.

Blinc, R. & Levanyuk, A. P. (1986). Editors. *Modern Problems in Condensed Matter Sciences, Incommensurate Phases in Dielectrics, 2. Materials.* Amsterdam: North-Holland.

Bouwman, W. G. & de Jeu, W. H. (1992). *3D XY behavior of a nematic–smectic-A phase transition: confirmation of the de Gennes model. Phys. Rev. Lett.* **68**, 800–803.

Brisbin, D., De Hoff, R., Lockhart, T. E. & Johnson, D. L. (1979). *Specific heat near the nematic–smectic-A tricritical point. Phys. Rev. Lett.* **43**, 1171–1174.

Brock, J. D., Aharony, A., Birgeneau, R. J., Evans-Lutterodt, K. W., Litster, J. D., Horn, P. M., Stephenson, G. B. & Tajbakhsh, A. R. (1986). *Orientational and positional order in a tilted hexatic liquid crystal phase. Phys. Rev. Lett.* **57**, 98–101.

Brooks, J. D. & Taylor, G. H. (1968). In *Chemistry and Physics of Carbon*, Vol. 3, edited by P. L. Walker Jr, pp. 243–286. New York: Marcel Dekker.

Bruinsma, R. & Nelson, D. R. (1981). *Bond orientational order in smectic liquid crystals. Phys. Rev. B*, **23**, 402–410.

Caillé, A. (1972). *Remarques sur la diffusion des rayons X dans les smectiques A. C. R. Acad. Sci. Sér. B*, **274**, 891–893.

Carlson, J. M. & Sethna, J. P. (1987). *Theory of the ripple phase in hydrated phospholipid bilayers. Phys. Rev. A*, **36**, 3359–3363.

Chan, K. K., Deutsch, M., Ocko, B. M., Pershan, P. S. & Sorensen, L. B. (1985). *Integrated X-ray scattering intensity measurement of the order parameter at the nematic to smectic-A phase transition. Phys. Rev. Lett.* **54**, 920–923.

Chan, K. K., Pershan, P. S., Sorensen, L. B. & Hardouin, F. (1985). *X-ray scattering study of the smectic-A1 to smectic-A2 transition. Phys. Rev. Lett.* **54**, 1694–1697.

Chan, K. K., Pershan, P. S., Sorensen, L. B. & Hardouin, F. (1986). *X-ray studies of transitions between nematic, smectic-A1, -A2, and -Ad phases. Phys. Rev. A*, **34**, 1420–1433.

Chandrasekhar, S. (1982). In *Advances in Liquid Crystals*, Vol. 5, edited by G. H. Brown, pp. 47–78. London, New York: Academic Press.

Chandrasekhar, S. (1983). *Liquid crystals of disk-like molecules. Philos. Trans. R. Soc. London Ser. A*, **309**, 93–103.

Chandrasekhar, S., Sadashiva, B. K. & Suresh, K. A. (1977). *Liquid crystals of disk like molecules. Pramana*, **9**, 471–480.

Chapman, D., Williams, R. M. & Ladbrooke, B. D. (1967). *Physical studies of phospholipids. VI. Thermotropic and lyotropic mesomorphism of some 1,2-diacylphosphatidylcholines (lecithins). Chem. Phys. Lipids*, **1**, 445–475.

Chen, J. H. & Lubensky, T. C. (1976). *Landau–Ginzberg mean-field theory for the nematic to smectic-C and nematic to smectic-A phase transitions. Phys. Rev. A*, **14**, 1202–1207.

Chu, K. C. & McMillan, W. L. (1977). *Unified Landau theory for the nematic, smectic A and smectic C phases of liquid crystals. Phys. Rev. A*, **15**, 1181–1187.

Coates, D. & Gray, G. W. (1975). *A correlation of optical features of amorphous liquid–cholesteric liquid crystal transitions. Phys. Lett. A*, **51**, 335–336.

Collett, J. (1983). PhD thesis, Harvard University, USA.

Collett, J., Pershan, P. S., Sirota, E. B. & Sorensen, L. B. (1984). *Synchrotron X-ray study of the thickness dependence of the phase diagram of thin liquid-crystal films. Phys. Rev. Lett.* **52**, 356–359.

Collett, J., Sorensen, L. B., Pershan, P. S. & Als-Nielsen, J. (1985). *X-ray scattering study of restacking transitions in the crystalline-B phases of heptyloxybenzylidene heptylaniline 7O.7. Phys. Rev. A*, **32**, 1036–1043.

Collett, J., Sorensen, L. B., Pershan, P. S., Litster, J., Birgeneau, R. J. & Als-Nielsen, J. (1982). *Synchrotron X-ray study of novel crystalline-B phases in 7O.7 Phys. Rev. Lett.* **49**, 553–556.

Crooker, P. P. (1983). *The cholesteric blue phase: a progress report. Mol. Cryst. Liq. Cryst.* **98**, 31–45.

Dasgupta, C. (1985). *Monte Carlo study of the nematic-to-smectic-A transition. Phys. Rev. Lett.* **55**, 1771–1774.

Dasgupta, C. (1987). *J. Phys. (Paris)*, **48**, 957–970.

Davey, S. C., Budai, J., Goodby, J. W., Pindak, R. & Moncton, D. E. (1984). *X-ray study of the hexatic-B to smectic-A phase transition in liquid crystal films. Phys. Rev. Lett.* **53**, 2129–2132.

Davidov, D., Safinya, C. R., Kaplan, M., Dana, S. S., Schaetzing, R., Birgeneau, R. J. & Litster, J. D. (1979). *High-resolution X-ray and light-scattering study of critical behavior associated with the nematic–smectic-A transition in 4-cyano-4′-octylbiphenyl. Phys. Rev. B*, **19**, 1657–1663.

De Gennes, P. G. (1969*a*). *Conjectures sur l'état smectique. J. Phys. (Paris) Colloq.* **30**, C4-65–C4-71.

De Gennes, P. G. (1969*b*). *Phenomenology of short-range-order effects in the isotropic phase of nematic materials. Phys. Lett. A*, **30**, 454–455.

De Gennes, P. G. (1971). *Short range order effects in the isotropic phase of nematics and cholesterics. Mol. Cryst. Liq. Cryst.* **12**, 193–214.

De Gennes, P. G. (1972). *An analogy between superconductors and smectics A. Solid State Commun.* **10**, 753–756.

De Gennes, P. G. (1973). *Some remarks on the polymorphism of smectics. Mol. Cryst. Liq. Cryst.* **21**, 49–76.

De Gennes, P. G. (1974). *The Physics of Liquid Crystals.* Oxford: Clarendon Press.

Demus, D., Diele, S., Klapperstück, M., Link, V. & Zaschke, H. (1971). *Investigation of a smectic tetramorphous substance. Mol. Cryst. Liq. Cryst.* **15**, 161–169.

Destrade, C., Gasparoux, H., Foucher, P., Tinh, N. H., Malthête, J. & Jacques, J. (1983). *Molecules discoides et polymorphisme mesomorphe. J. Chim. Phys.* **80**, 137–148.

De Vries, H. L. (1951). *Rotary power and other optical properties of liquid crystals. Acta Cryst.* **4**, 219–226.

Diele, S., Brand, P. & Sackmann, H. (1972). *X-ray diffraction and polymorphism of smectic liquid crystals. II. D and E modifications. Mol. Cryst. Liq. Cryst.* **17**, 163–169.

Djurek, D., Baturic-Rubcic, J. & Franulovic, K. (1974). *Specific-heat critical exponents near the nematic–smectic-A phase transition. Phys. Rev. Lett.* **33**, 1126–1129.

Doucet, J. (1979). In *Molecular Physics of Liquid Crystals*, edited by G. W. Gray & G. R. Luckhurst, pp. 317–341. London, New York: Academic Press.

Doucet, J. & Levelut, A. M. (1977). *X-ray study of the ordered smectic phases in some benzylideneanilines. J. Phys. (Paris)*, **38**, 1163–1170.

Doucet, J., Levelut, A. M., Lambert, M., Lievert, L. & Strzelecki, L. (1975). *Nature de la phase smectique. J. Phys. (Paris) Colloq.* **36**, C1-13–C1-19.

Etherington, G., Leadbetter, A. J., Wang, X. J., Gray, G. W. & Tajbakhsh, A. (1986). *Structure of the smectic D phase. Liq. Cryst.* **1**, 209–214.

Faetti, S., Gatti, M., Palleschi, V. & Sluckin, T. J. (1985). *Almost critical behavior of the anchoring energy at the interface between a nematic liquid crystal and a substrate. Phys. Rev. Lett.* **55**, 1681–1684.

Faetti, S. & Palleschi, V. (1984). *Nematic isotropic interface of some members of the homologous series of the 4-cyano-4′-(n-alkyl)biphenyl liquid crystals. Phys. Rev. A*, **30**, 3241–3251.

Fan, C. P. & Stephen, M. J. (1970). *Isotropic–nematic phase transition in liquid crystals. Phys. Rev. Lett.* **25**, 500–503.

Farber, A. S. (1985). PhD thesis, Brandeis University, USA.

Fleming, R. M., Moncton, D. E., McWhan, D. B. & DiSalvo, F. J. (1980). In *Ordering in Two Dimensions. Proceedings of an International Conference held at Lake Geneva, Wisconsin*, edited by S. K. Sinha, pp. 131–134. New York: North-Holland.

Fontell, K. (1974). In *Liquid Crystals and Plastic Crystals*, Vol. II, edited by G. W. Gray & P. A. Winsor, pp. 80–109. Chichester: Ellis Horwood.

Fontes, E., Heiney, P. A., Haseltine, J. H. & Smith, A. B. III (1986). *High resolution X-ray scattering study of the multiply reentrant polar mesogen DB9ONO$_2$. J. Phys. (Paris)*, **47**, 1533–1539.

Frank, F. C. & van der Merwe, J. H. (1949). *One-dimensional dislocations. I. Static theory. Proc. R. Soc. London Ser. A*, **198**, 205–216.

Freiser, M. J. (1971). *Successive transitions in a nematic liquid. Mol. Cryst. Liq. Cryst.* **14**, 165–182.

Friedel, G. (1922). *Les états mésomorphes de la matière. Ann. Phys. (Paris)*, **18**, 273–274.

Gane, P. A. C. & Leadbetter, A. J. (1981). *The crystal and molecular structure of N-(4-n-octyloxybenzylidene)-4′-butylaniline (8O.4) and the crystal–smectic G transition. Mol. Cryst. Liq. Cryst.* **78**, 183–200.

Gane, P. A. C. & Leadbetter, A. J. (1983). *Modulated crystal B phases and the B- to G-phase transition in two types of liquid crystalline compounds. J. Phys. C*, **16**, 2059–2067.

Gane, P. A. C., Leadbetter, A. J., Wrighton, P. G., Goodby, J. W., Gray, G. W. & Tajbakhsh, A. R. (1983). *The phase behavior of bis-(4′-n-heptyloxybenzylidene)-1,4-phenylenediamine HEPTOBPD, crystal J and K phases. Mol. Cryst. Liq. Cryst.* **100**, 67–74.

Gannon, M. G. J. & Faber, T. E. (1978). *The surface tension of nematic liquid crystals. Philos. Mag.* A**37**, 117–135.

Garland, C. W., Meichle, M., Ocko, B. M., Kortan, A. R., Safinya, C. R., Yu, L. J., Litster, J. D. & Birgeneau, R. J. (1983). *Critical behavior at the nematic–smectic-A transition in butyloxybenzylidene heptylaniline (4O.7). Phys. Rev. A*, **27**, 3234–3240.

Gransbergen, E. F., De Jeu, W. H. & Als-Nielsen, J. (1986). *Antiferroelectric surface layers in a liquid crystal as observed by synchrotron X-ray scattering. J. Phys. (Paris)*, **47**, 711–718.

Gray, G. W. & Goodby, J. W. (1984). *Smectic Liquid Crystals: Textures and Structures.* Glasgow: Leonard Hill.

Gray, G. W., Jones, B. & Marson, F. (1957). *Mesomorphism and chemical constitution. Part VIII. The effect of 3′-substituents on the mesomorphism of the 4′-n-alkoxydiphenyl-4-carboxylic acids and their alkyl esters. J. Chem. Soc.* **1**, 393–401.

Grinstein, G. & Toner, J. (1983). *Dislocation-loop theory of the nematic–smectic A–smectic C multicritical point. Phys. Rev. Lett.* **51**, 2386–2389.

Guillon, D. & Skoulios, A. (1987). *Molecular model for the R smectic DS mesophase. Europhys. Lett.* **3**, 79–85.

Guillon, D., Skoulios, A. & Benattar, J. J. (1986). *Volume and X-ray diffraction study of terephthal-bis-4-n-decylaniline (TBDA). J. Phys. (Paris)*, **47**, 133–138.

Gunther, L., Imry, Y. & Lajzerowicz, J. (1980). *X-ray scattering in smectic-A liquid crystals. Phys. Rev. A*, **22**, 1733–1740.

Guyot-Sionnest, P., Hsiung, H. & Shen, Y. R. (1986). *Surface polar ordering in a liquid crystal observed by optical second-harmonic generation. Phys. Rev. Lett.* **57**, 2963–2966.

Halperin, B. I. & Nelson, D. R. (1978). *Theory of two-dimensional melting. Phys. Rev. Lett.* **41**, 121–124, 519(E).

Hardouin, F., Levelut, A. M., Achard, M. F. & Sigaud, G. (1983). *Polymorphisme des substances mésogenes à molécules polaires. I. Physico-chimie et structure. J. Chim. Phys.* **80**, 53–64.

Hardouin, F., Levelut, A. M., Benattar, J. J. & Sigaud, G. (1980). *X-ray investigations of the smectic A1–smectic A2 transition. Solid State Commun.* **33**, 337–340.

Hardouin, F., Sigaud, G., Tinh, N. H. & Achard, M. F. (1981). *A fluid smectic A antiphase in a pure nitro rod-like compound. J. Phys. (Paris) Lett.* **42**, L63–L66.

Hardouin, F., Tinh, N. H., Achard, M. F. & Levelut, A. M. (1982). *A new thermotropic smectic phase made of ribbons. J. Phys. (Paris) Lett.* **43**, L327–L331.

Helfrich, W. (1979). *Structure of liquid crystals especially order in two dimensions. J. Phys. (Paris) Colloq.* **40**, C3-105–C3-114.

Hendrikx, Y., Charvolin, J. & Rawiso, M. (1986). *Uniaxial–biaxial transition in lyotropic nematic solutions: local biaxiality in the uniaxial phase. Phys. Rev. B*, **33**, 3534–3537.

Hirth, J. P., Pershan, P. S., Collett, J., Sirota, E. & Sorensen, L. B. (1984). *Dislocation model for restacking phase transitions in crystalline-B liquid crystals. Phys. Rev. Lett.* **53**, 473–476.

Hornreich, R. M., Luban, M. & Shtrikman, S. (1975). *Critical behavior at the onset of k-space instability on the lambda line. Phys. Rev. Lett.* **35**, 1678–1681.

Hornreich, R. M. & Shtrikman, S. (1983). *Theory of light scattering in cholesteric blue phases. Phys. Rev. A*, **28**, 1791–1807.

Huang, C. C. & Lien, S. C. (1981). *Nature of a nematic–smectic-A–smectic-C multicritical point. Phys. Rev. Lett.* **47**, 1917–1920.

Huang, C. C., Lien, S. C., Dumrongrattana, S. & Chiang, L. Y. (1984). *Calorimetric studies near the smectic-A1–smectic-$\tilde{A}$ phase transition of a liquid crystal compound. Phys. Rev. A*, **30**, 965–967.

Huse, D. A. (1985). *Fisher renormalization at the smectic-A1 to smectic-A2 transition in a mixture. Phys. Rev. Lett.* **55**, 2228.

Janiak, M. J., Small, D. M. & Shipley, G. G. (1979). *Temperature and compositional dependence of the structure of hydrated dimyristoyl lecithin. J. Biol. Chem.* **254**, 6068–6078.

Kasting, G. B., Lushington, K. J. & Garland, C. W. (1980). *Critical heat capacity near the nematic–smectic-A transition in octyloxycyanobiphenyl in the range 1–2000 bar. Phys. Rev. B*, **22**, 321–331.

Kittel, C. (1963). *Quantum Theory of Solids.* New York: John Wiley.

Kosterlitz, J. M. & Thouless, D. G. (1973). *Ordering, metastability and phase transitions in two-dimensions. J. Phys. C*, **6**, 1181–1203.

Landau, L. D. (1965). In *Collected Papers of L. D. Landau*, edited by D. ter Haar, pp. 193–216. New York: Gordon and Breach.

Landau, L. D. & Lifshitz, E. M. (1958). *Statistical Physics.* London: Pergamon Press.

Lawson, K. D. & Flautt, T. J. (1967). *Magnetically oriented lyotropic liquid crystalline phases. J. Am. Chem. Soc.* **89**, 5489–5491.

Leadbetter, A. J., Durrant, J. L. A. & Rugman, M. (1977). *The density of 4-n-octyl-4′-cyano-biphenyl (8CB). Mol. Cryst. Liq. Cryst.* **34**, 231–235.

Leadbetter, A. J., Frost, J. C., Gaughan, J. P., Gray, G. W. & Mosley, A. (1979). *The structure of smectic A phases of compounds with cyano end groups. J. Phys. (Paris) Colloq.* **40**, C3-375–C3-380.

Leadbetter, A. J., Frost, J. C., Gaughan, J. P. & Mazid, M. A. (1979). *The structure of the crystal, smectic E and smectic B forms of IBPAC. J. Phys. (Paris) Colloq.* **40**, C3-185–C3-192.

Leadbetter, A. J., Frost, J. C. & Mazid, M. A. (1979). *Interlayer correlations in smectic B phases. J. Phys. (Paris) Lett.* **40**, L325–L329.

Leadbetter, A. J., Gaughan, J. P., Kelley, B., Gray, G. W. & Goodby, J. W. (1979). *Characterisation and structure of some new smectic F phases. J. Phys. (Paris) Colloq.* **40**, C3-178–C3-184.

Leadbetter, A. J., Mazid, M. A. & Kelly, B. A. (1979). *Structure of the smectic-B phase and the nature of the smectic-B to H transition in the N-(4-n-alkoxybenzylidene)-4′-alkylanilines. Phys. Rev. Lett.* **43**, 630–633.

Leadbetter, A. J., Mazid, M. A. & Malik, K. M. A. (1980). *The crystal and molecular structure of isobutyl 4-(4′-phenylbenzylideneamino) cinnamate (IBPBAC) – and the crystal smectic E transition. Mol. Cryst. Liq. Cryst.* **61**, 39–60.

Leadbetter, A. J., Mazid, M. A. & Richardson, R. M. (1980). In *Liquid Crystals*, edited by S. Chandrasekhar, pp. 65–79. London: Heyden.

Leadbetter, A. J., Richardson, R. M. & Carlile, C. J. (1976). *The nature of the smectic E phase. J. Phys. (Paris) Colloq.* **37**, C3-65–C3-68.

Lee, S. D. & Meyer, R. B. (1986). *Computations of the phase equilibrium, elastic constants, and viscosities of a hard-rod nematic liquid crystal. J. Chem. Phys.* **84**, 3443–3448.

Le Guillou, J. C. & Zinn-Justin, J. (1985). *J. Phys. Lett.* **52**, L-137–L-141.

Levelut, A. M. (1976). *Etude de l'ordre local lié à la rotation des molécules dans la phase smectique B. J. Phys.* (*Paris*) *Colloq.* **37**, C3-51–C3-54.

Levelut, A. M. (1983). *Structures des phases mésomorphes formée de molécules discoides. J. Chim. Phys.* **80**, 149–161.

Levelut, A. M., Doucet, J. & Lambert, M. (1974). *Etude par diffusion de rayons X de la nature des phases smectiques B et de la transition de phase solide–smectique B. J. Phys. (Paris)*, **35**, 773–779.

Levelut, A. M. & Lambert, M. (1971). *Structure des cristaux liquides smectic B. C. R. Acad Sci. Sér. B*, **272**, 1018–1021.

Levelut, A. M., Tarento, R. J., Hardouin, F., Achard, M. F. & Sigaud, G. (1981). *Number of SA phases. Phys. Rev. A*, **24**, 2180–2186.

Liebert, L. E. (1978). Editor. *Liquid Crystals.* In *Solid State Physics: Advances in Research and Applications*, edited by H. Ehrenreich, F. Seitz & D. Turnbull, Suppl. 14. New York: Academic Press.

Litster, J. D., Als-Nielsen, J., Birgeneau, R. J., Dana, S. S., Davidov, D., Garcia-Golding, F., Kaplan, M., Safinya, C. R. & Schaetzing, R. (1979). *High resolution X-ray and light scattering studies of bilayer smectic A compounds. J. Phys. (Paris) Colloq.* **40**, C3-339–C3-344.

Lubensky, T. C. (1983). *The nematic to smectic A transition: a theoretical overview. J. Chim. Phys.* **80**, 31–43.

Lubensky, T. C. (1987). *Mean field theory for the biaxial nematic phase and the NNUAC critical point. Mol. Cryst. Liq. Cryst.* **146**, 55–69.

Lubensky, T. C., Ramaswamy, S. & Toner, J. (1988). *Static and dynamic properties of incommensurate smectic-A_{IC} liquid crystals. Phys. Rev. A*, **38**, 4284–4298.

Luzzati, V. (1968). In *Biological Membranes: Physical Fact and Function*, Vol. 1, edited by D. Chapman, pp. 71–123. New York: Academic Press.

Luzzati, V. & Reiss-Husson, F. (1966). *Structure of the cubic phase of lipid–water systems. Nature (London)*, **210**, 1351–1352.

Ma, S. K. (1976). *Modern Theory of Critical Phenomena.* Reading: Benjamin.

McMillan, W. L. (1972). *X-ray scattering from liquid crystals. I. Cholesteryl nonanoate and myristate. Phys. Rev. A*, **6**, 936–947.

McMillan, W. L. (1973*a*) *Measurement of smectic-A-phase order-parameter fluctuations near a second-order smectic-A–nematic-phase transition. Phys. Rev. A*, **7**, 1419–1422.

McMillan, W. L. (1973*b*). *Measurement of smectic-A-phase order-parameter fluctuations in the nematic phase of p-n-octyloxybenzylidene-p′-toluidine. Phys. Rev. A*, **7**, 1673–1678.

McMillan, W. L. (1973*c*). *Measurement of smectic-phase order-parameter fluctuations in the nematic phase of heptyloxybenzene. Phys. Rev. A*, **8**, 328–331.

Mada, H. & Kobayashi, S. (1981). *Surface order parameter of 4-n-heptyl-4′-cyanobiphenyl. Mol. Cryst. Liq. Cryst.* **66**, 57–60.

Maier, W. & Saupe, A. (1958). *Eine einfache molekulare Theorie des nematischen kristallinglüssigen Zustandes. Z. Naturforsch. Teil A*, **13**, 564–566.

Maier, W. & Saupe, A. (1959). *Eine einfache molekularestatistiche Theorie der nematischen kristallinglüssigen phase. Teil I. Z. Naturforsch. Teil A*, **14**, 882–889.

Malthête, J., Liebert, L., Levelut, A. M. & Galerne, Y. (1986). *Nématic biaxe thermotrope. C. R. Acad. Sci.* **303**, 1073–1076.

Martin, P. C., Parodi, O. & Pershan, P. S. (1972). *Unified hydrodynamic theory for crystals, liquid crystals, and normal fluids. Phys. Rev. A*, **6**, 2401–2420.

Martinez-Miranda, L. J., Kortan, A. R. & Birgeneau, R. J. (1986). *X-ray study of fluctuations near the nematic–smectic-A–smectic-C multicritical point. Phys. Rev. Lett.* **56**, 2264–2267.

Meiboom, S., Sethna, J. P., Anderson, P. W. & Brinkman, W. F. (1981). *Theory of the blue phase of cholesteric liquid crystals. Phys. Rev. Lett.* **46**, 1216–1219.

Miyano, K. (1979). *Wall-induced pretransitional birefringence: a new tool to study boundary aligning forces in liquid crystals. Phys. Rev. Lett.* **43**, 51–54.

Moncton, D. E. & Pindak, R. (1979). *Long-range order in two- and three-dimensional smectic-B liquid crystal films. Phys. Rev. Lett.* **43**, 701–704.

Moncton, D. E., Pindak, R., Davey, S. C. & Brown, G. S. (1982). *Melting of variable thickness liquid crystal thin films: a synchrotron X-ray study. Phys. Rev. Lett.* **49**, 1865–1868.

Moncton, D. E., Stephens, P. W., Birgeneau, R. J., Horn, P. M. & Brown, G. S. (1981). *Synchrotron X-ray study of the commensurate–incommensurate transition of monolayer krypton on graphite. Phys. Rev. Lett.* **46**, 1533–1536.

Nelson, D. R. & Halperin, B. I. (1979). *Dislocation-mediated melting in two dimensions. Phys. Rev. B*, **19**, 2457–2484.

Nelson, D. R. & Halperin, B. I. (1980). *Solid and fluid phases in smectic layers with tilted molecules. Phys. Rev. B*, **21**, 5312–5329.

Nelson, D. R. & Toner, J. (1981). *Bond-orientational order, dislocation loops and melting of solids and smectic-A liquid crystals. Phys. Rev. B*, **24**, 363–387.

Neto, A. M. F., Galerne, Y., Levelut, A. M. & Liebert, L. (1985). *Pseudo lamellar ordering in uniaxial and biaxial lyotropic nematics: a synchrotron X-ray diffraction experiment. J. Phys. (Paris) Lett.* **46**, L499–L505.

Ocko, B. M., Birgeneau, R. J. & Litster, J. D. (1986). *Crossover to tricritical behavior at the nematic to smectic A transition: an X-ray scattering study. Z. Phys.* B**62**, 487–497.

Ocko, B. M., Birgeneau, R. J., Litster, J. D. & Neubert, M. E. (1984). *Critical and tricritical behavior at the nematic to smectic-A transition. Phys. Rev. Lett.* **52**, 208–211.

Ocko, B. M., Braslau, A., Pershan, P. S., Als-Nielsen, J. & Deutsch, M. (1986). *Quantized layer growth at liquid crystal surfaces. Phys. Rev. Lett.* **57**, 94–97.

Ocko, B. M., Pershan, P. S., Safinya, C. R. & Chiang, L. Y. (1987). *Incommensurate smectic order at the free surface in the nematic phase of $DB7NO_2$. Phys. Rev. A*, **35**, 1868–1872.

Onsager, L. (1949). *The effects of shapes on the interaction of colloidal particles. Ann. NY Acad. Sci.* **51**, 627–659.

Peierls, R. E. (1934). *Transformation temperatures. Helv. Phys. Acta Suppl.* **2**, 81–83.

Pershan, P. S. (1974). *Dislocation effects in smectic-A liquid crystals. J. Appl. Phys.* **45**, 1590–1604.

Pershan, P. S. (1979). *Amphiphilic molecules and liquid crystals. J. Phys. (Paris) Colloq.* **40**, C3-423–C3-432.

Pershan, P. S. & Als-Nielsen, J. (1984). *X-ray reflectivity from the surface of a liquid crystal: surface structure and absolute value of critical fluctuations. Phys. Rev. Lett.* **52**, 759–762.

Pershan, P. S., Braslau, A., Weiss, A. H. & Als-Nielsen, J. (1987). *Smectic layering at the free surface of liquid crystals in the nematic phase: X-ray reflectivity. Phys. Rev. A*, **35**, 4800–4813.

Pershan, P. S. & Prost, J. (1975). *Dislocation and impurity effects in smectic-A liquid crystals. J. Appl. Phys.* **46**, 2343–2353.

Pindak, R., Moncton, D. E., Davey, S. C. & Goodby, J. W. (1981). *X-ray observation of a stacked hexatic liquid-crystal B phase. Phys. Rev. Lett.* **46**, 1135–1138.

Pokrovsky, V. L. & Talapov, A. L. (1979). *Ground state, spectrum, and phase diagram of two-dimensional incommensurate crystals. Phys. Rev. Lett.* **42**, 65–67.

Prost, J. (1979). *Smectic A to smectic A phase transition. J. Phys. (Paris)*, **40**, 581–587.

Prost, J. (1984). *The smectic state. Adv. Phys.* **33**, 1–46.

Prost, J. & Barois, P. (1983). *Polymorphism in polar mesogens. II. Theoretical aspects. J. Chim. Phys.* **80**, 65–81.

Ratna, B. R., Nagabhushana, C., Raja, V. N., Shashidhar, R. & Chandrasekhar, S. (1986). *Density, dielectric and X-ray studies of smectic A–smectic A transitions. Mol. Cryst. Liq. Cryst.* **138**, 245–257.

Ratna, B. R., Shashidhar, R. & Raja, V. N. (1985). *Smectic-A phase with two collinear incommensurate density modulations. Phys. Rev. Lett.* **55**, 1476–1478.

Richardson, R. M., Leadbetter, A. J. & Frost, J. C. (1978). *The structure and dynamics of the smectic B phase. Ann. Phys.* **3**, 177–186.

Sadoc, J. F. & Charvolin, J. (1986). *Frustration in bilayers and topologies of liquid crystals of amphiphilic molecules. J. Phys. (Paris)*, **47**, 683–691.

Safinya, C. R., Clark, N. A., Liang, K. S., Varady, W. A. & Chiang, L. Y. (1985). *Synchrotron X-ray scattering study of freely suspended discotic strands. Mol. Cryst. Liq. Cryst.* **123**, 205–216.

Safinya, C. R., Liang, K. S., Varady, W. A., Clark, N. A. & Andersson, G. (1984). *Synchrotron X-ray study of the orientational ordering D2–D1 structural phase transition of freely suspended discotic strands in triphenylene hexa-n-dodecanoate. Phys. Rev. Lett.* **53**, 1172–1175.

Safinya, C. R., Roux, D., Smith, G. S., Sinha, S. K., Dimon, P., Clark, N. A. & Bellocq, A. M. (1986). *Steric interactions in a model multimembrane system: a synchrotron X-ray study. Phys. Rev. Lett.* **57**, 2718–2721.

Safinya, C. R., Varady, W. A., Chiang, L. Y. & Dimon, P. (1986). *X-ray study of the nematic phase and smectic-A1 to smectic-$\tilde{A}$ phase transition in heptylphenyl nitrobenzoyloxybenzoate (DB7NO2). Phys. Rev. Lett.* **57**, 432–435.

Safran, S. A. & Clark, N. A. (1987). Editors. *Physics of Complex and Supermolecular Fluids.* New York: John Wiley.

Selinger, J. V. & Nelson, D. R. (1988). *Theory of hexatic-to-hexatic transitions*. *Phys. Rev. Lett.* **61**, 416–419.

Shipley, G. G., Hitchcock, P. B., Mason, R. & Thomas, K. M. (1974). *Structural chemistry of 1,2-diauroyl-*DL*-phosphatidylethanolamine: molecular conformation and intermolecular packing of phospholipids. Proc. Natl Acad. Sci. USA*, **71**, 3036–3040.

Sigaud, G., Hardouin, F., Achard, M. F. & Gasparoux, H. (1979). *Anomalous transitional behaviour in mixtures of liquid crystals: a new transition of SA–SA type? J. Phys. (Paris) Colloq.* **40**, C3-356–C3-359.

Sirota, E. B., Pershan, P. S. & Deutsch, M. (1987). *Modulated crystalline-B phases in liquid crystals. Phys. Rev. A*, **36**, 2902–2913.

Sirota, E. B., Pershan, P. S., Sorensen, L. B. & Collett, J. (1985). *X-ray studies of tilted hexatic phases in thin liquid-crystal films. Phys. Rev. Lett.* **55**, 2039–2042.

Sirota, E. B., Pershan, P. S., Sorensen, L. B. & Collett, J. (1987). *X-ray and optical studies of the thickness dependence of the phase diagram of liquid crystal films. Phys. Rev. A*, **36**, 2890–2901.

Sirota, E. B., Smith, G. S., Safinya, C. R., Plano, R. J. & Clark, N. A. (1988). *X-ray scattering studies of aligned, stacked surfactant membranes*. *Science*, **242**, 1406–1409.

Small, D. (1967). *Phase equilibria and structure of dry and hydrated egg lecithin. J. Lipid Res.* **8**, 551–557.

Smith, G. W. & Garland, Z. G. (1973). *Liquid crystalline phases in a doubly homologous series of benzylideneanilines – textures and scanning calorimetry. J. Chem. Phys.* **59**, 3214–3228.

Smith, G. W., Garland, Z. G. & Curtis, R. J. (1973). *Phase transitions in mesomorphic benzylideneanilines. Mol. Cryst. Liq. Cryst.* **19**, 327–330.

Smith, G. S., Sirota, E. B., Safinya, C. R. & Clark, N. A. (1988). *Structure of the L_β phases in a hydrated phosphatidylcholine multimembrane*. *Phys. Rev. Lett.* **60**, 813–816.

Solomon, L. & Litster, J. D. (1986). *Light scattering measurements in the 7S5-8OCB nematic–smectic-A–smectic-C liquid-crystal system. Phys. Rev. Lett.* **56**, 2268–2271.

Sorensen, L. B. (1987). Private communication.

Sorensen, L. B., Amador, S., Sirota, E. B., Stragler, H. & Pershan, P. S. (1987). Unpublished.

Springer, T. (1977). In *Current Physics*, Vol. 3. *Dynamics of Solids and Liquids by Neutron Scattering*, edited by S. W. Lovesey & T. Springer, pp. 255–300 (see p. 281). Berlin: Springer-Verlag.

Sprokel, G. E. (1980). *The Physics and Chemistry of Liquid Crystal Devices.* New York: Plenum.

Stegemeyer, H. & Bergmann, K. (1981). In *Liquid Crystals of One- and Two-Dimensional Order. Springer Series in Chemical Physics* 11, edited by W. Helfrich & G. Heppke, pp. 161–175. Berlin, Heidelberg, New York: Springer-Verlag.

Tardieu, A. & Billard, J. (1976). *On the structure of the 'smectic D modification'. J. Phys. (Paris) Colloq.* **37**, C3-79–C3-81.

Tardieu, A. & Luzzati, V. (1970). *Polymorphism of lipids: a novel cubic phase – a cage-like network of rods with enclosed spherical micelles. Biochim. Biophys. Acta*, **219**, 11–17.

Tardieu, A., Luzzati, V. & Reman, F. C. (1973). *Structure and polymorphism of the hydrocarbon chains of lipids: a study of lecithin–water phases*. *J. Mol. Biol.* **75**, 711–733.

Thoen, J., Marynissen, H. & Van Dael, W. (1982). *Temperature dependence of the enthalpy and the heat capacity of the liquid-crystal octylcyanobiphenyl (8CB). Phys. Rev. A*, **26**, 2886–2905.

Thoen, J., Marynissen, H. & Van Dael, W. (1984). *Nematic–smectic-A tricritical point in alkylcyanobiphenyl liquid crystals. Phys. Rev. Lett.* **5**, 204–207.

Villain, J. (1980). In *Order in Strongly Fluctuating Condensed Matter System*, edited by T. Riste, pp. 221–260. New York: Plenum.

Wang, J. & Lubensky, T. C. (1984). *Theory of the SA1–SA2 phase transition in liquid crystal. Phys. Rev. A*, **29**, 2210–2217.

Warren, B. E. (1968). *X-ray Diffraction.* Reading: Addison-Wesley.

Winkor, M. J. & Clarke, R. (1986). *Long-period stacking transitions in intercalated graphite. Phys. Rev. Lett.* **56**, 2072–2075.

Young, A. P. (1979). *Melting and the vector Coulomb gas in two dimensions. Phys. Rev. B*, **19**, 1855–1866.

Young, C. Y., Pindak, R., Clark, N. A. & Meyer, R. B. (1978). *Light-scattering study of two-dimensional molecular-orientation fluctuations in a freely suspended ferroelectric liquid crystal film. Phys. Rev. Lett.* **40**, 773–776.

Yu, L. J. & Saupe, A. (1980). *Observation of a biaxial nematic phase in potassium laurate–1-decanol–water mixtures. Phys. Rev. Lett.* **45**, 1000–1003.

International Tables for Crystallography (2010). Vol. B, Chapter 4.5, pp. 567–589.

4.5. Polymer crystallography

By R. P. Millane and D. L. Dorset

4.5.1. Overview

By R. P. Millane and D. L. Dorset

Linear polymers from natural or synthetic sources are actually polydisperse aggregates of high-molecular-weight chains. Nevertheless, many of these essentially infinite-length molecules can be prepared as solid-state specimens that contain ordered molecular segments or crystalline inclusions (Vainshtein, 1966; Tadokoro, 1979; Mandelkern, 1989; Barham, 1993). In general, ordering can occur in a number of ways. Hence an oriented and/or somewhat ordered packing of chain segments might be found in a stretched fibre, or in the chain-folded arrangement of a lamellar crystallite. Lamellae themselves may exist as single plates or in the more complex array of a spherulite (Geil, 1963). Diffraction data can be obtained from these various kinds of specimens and used to determine molecular and crystal structures.

There are numerous reasons why crystallography of polymers is important. Although it may be possible to crystallize small constituent fragments of these large molecules and determine their crystal structures, one often wishes to study the intact (and biologically or functionally active) polymeric system. The molecular conformations and intermolecular interactions are determinants of parameters such as persistence length which affect, for example, solution conformations (random or worm-like coils) which determine viscosity. Molecular conformations also influence intermolecular interactions, which determine physical properties in gels and melts. Molecular conformations are, of course, of critical importance in many biological recognition processes. Knowledge of the stereochemical constraints that are placed on the molecular packing to maximize unit-cell density is particularly relevant to the fact that many linear molecules (as well as monodisperse substances with low molecular weight) can adopt several different allomorphic forms, depending on the crystallization conditions employed or the biological origin. Since different allomorphs can behave quite differently from one another, it is clear that the mode of chain packing is related to the bulk properties of the polymer (Grubb, 1993). The three-dimensional geometry of the chain packing obtained from a crystal structure analysis can be used to investigate other phenomena such as the possible inclusion of disordered material in *chain-fold* regions (Mandelkern, 1989; Lotz & Wittmann, 1993), the ordered interaction of crystallite sectors across grain boundaries where tight interactions are found between domains, or the specific interactions of polymer chains with another substance in a composite material (Lotz & Wittmann, 1993).

The two primary crystallographic techniques used for studying polymer structure are described in this chapter. The first is X-ray fibre diffraction analysis, described in Section 4.5.2; and the second is polymer electron crystallography, described in Section 4.5.3.

Crystallographic studies of polymers were first performed using X-ray diffraction from oriented fibre specimens. Early applications were to cellulose and DNA from the 1930s to the 1950s, and the technique has subsequently been applied to hundreds of biological and synthetic polymers (Arnott, 1980; Millane, 1988). This technique is now referred to as *X-ray fibre diffraction analysis*. In fact, fibre diffraction analysis can be employed not only for polymers, but for any system that can be oriented. Indeed, one of the first applications of the technique was to tobacco mosaic virus (Franklin, 1955). Fibre diffraction analysis has also utilized, in some cases, neutrons instead of X-rays (*e.g.* Stark *et al.*, 1988; Forsyth *et al.*, 1989). X-ray fibre diffraction analysis is particularly suitable for biological polymers that form natural fibrous superstructures and even for many synthetic polymers that exist in either a fibrous or a liquid-crystalline state. Fibre diffraction has played an important role in structural studies of polynucleotides, polysaccharides, polypeptides and polyesters, as well as rod-like helical viruses, bacteriophages, microtubules and muscle fibres (Arnott, 1980; French & Gardner, 1980; Hall, 1984; Millane, 1988; Atkins, 1989). The common, and unique, feature of these systems is that the molecules (or their aggregates) are randomly rotated about an axis of preferred orientation. As a result, the measured diffraction is the cylindrical average of that from a single molecule or aggregate. The challenge for the structural scientist, therefore, is that of structure determination from cylindrically averaged diffraction intensities. Since a wide range of types and degrees of order (or disorder) occur in fibrous specimens, as well as a wide range of sizes of the repeating units, a variety of methods are used for structure determination.

The second technique used for structural studies of polymers is *polymer electron crystallography*. This involves measuring electron intensity data from individual crystalline regions or lamellae in the diffraction plane of an electron microscope. This is possible because a narrow electron beam can be focused on a single thin microcrystal and because of the enhanced scattering cross section of matter for electrons. By tilting the specimen, three-dimensional diffraction intensities from a single microcrystal can be collected. This means that the unit-cell dimensions and symmetry can be obtained unambiguously in electron-diffraction experiments on individual chain-folded lamellae, and the data can be used for actual single-crystal structure determinations. One of the first informative electron-diffraction studies of crystalline polymer films was made by Storks (1938), who formulated the concept of chain folding in polymer lamellae. Among the first quantitative structure determinations from electron-diffraction intensities was that of Tatarinova & Vainshtein (1962) on the α form of poly-γ-methyl-L-glutamate. Quantitative interpretation of the intensity data may benefit from the assumption of *quasi-kinematical* scattering (Dorset, 1995*a*), as long as the proper constraints are placed on the experiment. Structure determination may then proceed using the traditional techniques of X-ray crystallography. While molecular-modelling approaches (in which atomic level molecular and crystal structure models are constructed and refined) have been employed with single-crystal electron-diffraction data (Brisse, 1989), the importance of *ab initio* structure determination has been established in recent years (Dorset, 1995*b*), demonstrating that no initial assumptions about the molecular geometry need be made before the determination is begun. In some cases too, high-resolution electron micrographs of the polymer crystal structure can be used as an additional means for determining crystallographic phases and/or to visualize lattice defects.

Each of the two techniques described above has its own advantages and disadvantages. While specimen disorder can limit the application of X-ray fibre diffraction analysis, polymer electron diffraction is limited to materials that can be be prepared as crystalline lamallae and that can withstand the high vacuum environment of an electron microscope (although the latter restriction can now be largely overcome by the use of low-temperature specimen holders and/or environmental chambers).

Copyright © 2010 International Union of Crystallography

4.5.2. X-ray fibre diffraction analysis

By R. P. Millane

4.5.2.1. Introduction

X-ray fibre diffraction analysis is a collection of crystallographic techniques that are used to determine molecular and crystal structures of molecules, or molecular assemblies, that form specimens (often fibres) in which the molecules, assemblies or crystallites are approximately parallel but not otherwise ordered (Arnott, 1980; French & Gardner, 1980; Hall, 1984; Vibert, 1987; Millane, 1988; Atkins, 1989; Stubbs, 1999). These are usually long, slender molecules and they are often inherently flexible, which usually precludes the formation of regular three-dimensional crystals suitable for conventional crystallographic analysis. X-ray fibre diffraction therefore provides a route for structure determination for certain kinds of specimens that cannot be crystallized. Although it may be possible to crystallize small *fragments* or *subunits* of these molecules, and determine the crystal structures of these, X-ray fibre diffraction provides a means for studying the intact, and often the biologically or functionally active, system. Fibre diffraction has played an important role in the determination of biopolymers such as polynucleotides, polysaccharides (both linear and branched), polypeptides and a wide variety of synthetic polymers (such as polyesters), as well as larger assemblies including rod-like helical viruses, bacteriophages, microtubules and muscle fibres (Arnott, 1980; Arnott & Mitra, 1984; Millane, 1990*c*; Squire & Vibert, 1987).

Specimens appropriate for fibre diffraction analysis exhibit rotational disorder (of the molecules, aggregates or crystallites) about a preferred axis, resulting in cylindrical averaging of the diffracted intensity in reciprocal space. Therefore, fibre diffraction analysis can be thought of as 'structure determination from cylindrically averaged diffraction intensities' (Millane, 1993). In a powder specimen the crystallites are completely (spherically) disordered, so that structure determination by fibre diffraction can be considered to be intermediate between structure determination from single crystals and from powders.

This section is a review of the theory and techniques of structure determination by X-ray fibre diffraction analysis. It includes descriptions of fibre specimens, the theory of diffraction by these specimens, intensity data collection and processing, and the variety of structure determination methods used for the various kinds of specimens studied by fibre diffraction. It does not include descriptions of specimen preparation (those can be found in the references given for specific systems), or of applications of X-ray diffraction to determining polymer morphology (*e.g.* particle or void sizes and shapes, texture, domain structure *etc.*).

4.5.2.2. Fibre specimens

A wide variety of kinds of fibre specimen exist. All exhibit preferred orientation; the variety results from variability in the degree of order (crystallinity) in the lateral plane (the plane perpendicular to the axis of preferred orientation). This leads to categorization of three kinds of fibre specimen: *noncrystalline fibres*, in which there is no order in the lateral plane; *polycrystalline fibres*, in which there is near-perfect crystallinity in the lateral plane; and *disordered fibres*, in which there is disorder either within the molecules or in their crystalline packing (or both). The kind of fibre specimen affects the kind of diffraction pattern obtained, the relationships between the molecular and crystal structures and the diffraction data, methods of data collection, and methods of structure determination.

Noncrystalline fibres are made up of a collection of molecules that are *oriented*. This means that there is a common axis in each molecule (referred to here as the *molecular axis*), the axes being parallel in the specimen. The direction of preferred orientation is called the *fibre axis*. The molecule itself is usually considered to be a rigid body. There is no other ordering within the specimen. The molecules are therefore randomly positioned in the lateral plane and are randomly rotated about their molecular axes. Furthermore, if the molecule does not have a twofold rotation axis normal to the molecular axis, then the molecular axis has a *direction* associated with it, and the molecular axes are oriented randomly parallel or antiparallel to each other. This is often called *directional disorder*, or the molecules are said to be oriented *randomly up and down*. The average length of the ordered molecular segments in a noncrystalline fibre is referred to as the *coherence length*.

Polycrystalline fibres are characterized by molecular segments packing together to form well ordered microcrystallites within the specimen. The crystallites effectively take the place of the molecules in a noncrystalline specimen as described above. The crystallites are oriented, and since the axis within each crystallite that is aligned parallel to those in other crystallites usually corresponds to the long axes of the constituent molecules, it is also referred to here as the molecular axis. The crystallites are randomly positioned in the lateral plane, randomly rotated about the molecular axis, and randomly oriented up or down. The size of the crystalline domains can be characterized by their average dimensions in the directions of the **a**, **b** and **c** unit-cell vectors. However, because of the rotational disorder of the crystallites, any differences between crystallite dimensions in different directions normal to the fibre axis tend to be smeared out in the diffraction pattern, and the crystallite size is usefully characterized by the average dimensions of the crystallites normal and parallel to the fibre axis.

The molecules or crystallites in a fibre specimen are not perfectly oriented, and the variation in inclinations of the molecular axes to the fibre axis is referred to as *disorientation*. Assuming that the orientation is axisymmetric, then it can be described by an *orientation density function* $\Omega(\alpha)$ such that $\Omega(\alpha)\,\mathrm{d}\omega$ is the fraction of molecules in an element of solid angle $\mathrm{d}\omega$ inclined at an angle α to the fibre axis. The exact form of $\Omega(\alpha)$ is generally not known for any particular fibre and it is often sufficient to assume a Gaussian orientation density function, so that

$$\Omega(\alpha) = \frac{1}{2\pi\alpha_0^2}\exp\left(-\frac{\alpha^2}{2\alpha_0^2}\right), \qquad (4.5.2.1)$$

where α_0 is a measure of the degree of disorientation.

Fibre specimens often exhibit various kinds of disorder. The disorder may be within the molecules or in their packing. Disorder affects the relationship between the molecular and crystal structure and the diffracted intensities. Disorder within the molecules may result from a degree of randomness in the chemical sequence of the molecule or from variability in the interactions between the units that make up the molecule. Such molecules may (at least in principle) form noncrystalline, polycrystalline or partially crystalline (described below) fibres. Disordered packing of molecules within crystallites can result from a variety of ways in which the molecules can interact with each other. Fibre specimens made up of disordered crystallites are referred to here as partially crystalline fibres.

4.5.2.3. Diffraction by helical structures

Molecules or assemblies studied by fibre diffraction are usually made up of a large number of identical, or nearly identical, residues, or subunits, that in an oriented specimen are distributed along an axis; this leads naturally to helical symmetry. Since a periodic structure with no helix symmetry can be treated as a onefold helix, the assumption of helix symmetry is not restrictive.

4.5.2.3.1. Helix symmetry

The presence of a unique axis about which there is rotational disorder means that it is convenient to use cylindrical polar coordinate systems in fibre diffraction. We denote by (r, φ, z) a cylindrical polar coordinate system in real space, in which the z axis is parallel to the molecular axes. The molecule is said to have u_v helix symmetry, where u and v are integers, if the electron density $f(r, \varphi, z)$ satisfies

$$f\big(r, \varphi + (2\pi mv/u), z + (mc/u)\big) = f(r, \varphi, z), \qquad (4.5.2.2)$$

where m is any integer. The constant c is the period along the z direction, which is referred to variously as the *molecular repeat distance*, the *crystallographic repeat*, or the *c repeat*. The helix *pitch* P is equal to c/v. Helix symmetry is easily interpreted as follows. There are u subunits, or *helix repeat units*, in one c repeat of the molecule. The helix repeat units are repeated by integral rotations of $2\pi v/u$ about, and translations of c/u along, the molecular (or helix) axis. The helix repeat units may therefore be referenced to a helical lattice that consists of points at a fixed radius, with relative rotations and translations as described above. These points lie on a helix of pitch P, there are v turns (or pitch-lengths) of the helix in one c repeat, and there are u helical lattice points in one c repeat. A u_v helix is said to have 'u residues in v turns'.

Since the electron density is periodic in φ and z, it can be decomposed into a Fourier series as

$$f(r, \varphi, z) = \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} g_{nl}(r) \exp\big(i[n\varphi - (2\pi lz/c)]\big), \qquad (4.5.2.3)$$

where the coefficients $g_{nl}(r)$ are given by

$$g_{nl}(r) = (c/2\pi) \int_0^c \int_0^{2\pi} f(r, \varphi, z) \exp\big(i[-n\varphi + (2\pi lz/c)]\big)\, \mathrm{d}\varphi\, \mathrm{d}z. \qquad (4.5.2.4)$$

Assume now that the electron density has helical symmetry. Denote by $g(r, \varphi, z)$ the electron density in the region $0 < z < c/u$; the electron density being zero outside this region, *i.e.* $g(r, \varphi, z)$ is the electron density of a single helix repeat unit. It follows that

$$f(r, \varphi, z) = \sum_{m=-\infty}^{\infty} g[r, \varphi + (2\pi mv/u), z + (mc/u)]. \qquad (4.5.2.5)$$

Substituting equation (4.5.2.5) into equation (4.5.2.4) shows that $g_{nl}(r)$ vanishes unless $(l - nv)$ is a multiple of u, *i.e.* unless

$$l = um + vn \qquad (4.5.2.6)$$

for any integer m. Equation (4.5.2.6) is called the *helix selection rule*. The electron density in the helix repeat unit is therefore given by

$$g(r, \varphi, z) = \sum_l \sum_n g_{nl}(r) \exp\big(i[n\varphi - (2\pi lz/c)]\big), \qquad (4.5.2.7)$$

where

$$g_{nl}(r) = (c/2\pi) \int\int g(r, \varphi, z) \exp\big(i[-n\varphi + (2\pi l/c)]\big)\, \mathrm{d}\varphi\, \mathrm{d}z, \qquad (4.5.2.8)$$

and where in equation (4.5.2.7) (and in the remainder of this section) the sum over l is over all integers, the sum over n is over all integers satisfying the helix selection rule and the integral in equation (4.5.2.8) is over one helix repeat unit. The effect of helix symmetry, therefore, is to restrict the number of Fourier coefficients $g_{nl}(r)$ required to represent the electron density to those whose index n satisfies the selection rule. Note that the selection rule is usually derived using a rather more complicated argument by considering the convolution of the Fourier transform of a continuous filamentary helix with a set of planes in reciprocal space (Cochran *et al.*, 1952). The approach described above, which follows that of Millane (1991), is much more straightforward.

4.5.2.3.2. Diffraction by helical structures

Denote by (R, ψ, Z) a cylindrical polar coordinate system in reciprocal space (with the Z and z axes parallel), and by $F(R, \psi, Z)$ the Fourier transform of $f(r, \varphi, z)$. Since $f(r, \varphi, z)$ is periodic in z with period c, its Fourier transform is nonzero only on the *layer planes* $Z = l/c$ where l is an integer. Denote $F(R, \psi, l/c)$ by $F_l(R, \psi)$; using the cylindrical form of the Fourier transform shows that

$$F_l(R, \psi) = \int_0^c \int_0^{2\pi} \int_0^{\infty} f(r, \varphi, z) \exp\big(i\, 2\pi\, [Rr\cos(\psi - \varphi) + (lz/c)]\big) r\, \mathrm{d}r\, \mathrm{d}\varphi\, \mathrm{d}z. \qquad (4.5.2.9)$$

It is convenient to rewrite equation (4.5.2.9) making use of the Fourier decomposition described in Section 4.5.2.3.1, since this allows utilization of the helix selection rule. The *Fourier–Bessel structure factors* (Klug *et al.*, 1958), $G_{nl}(R)$, are defined as the Hankel transform of the Fourier coefficients $g_{nl}(r)$, *i.e.*

$$G_{nl}(R) = \int_0^{\infty} g_{nl}(r) J_n(2\pi Rr) 2\pi r\, \mathrm{d}r, \qquad (4.5.2.10)$$

and the inverse transform is

$$g_{nl}(r) = \int_0^{\infty} G_{nl}(R) J_n(2\pi Rr) 2\pi R\, \mathrm{d}R. \qquad (4.5.2.11)$$

Using equations (4.5.2.7) and (4.5.2.11) shows that equation (4.5.2.9) can be written as

$$F_l(R, \psi) = \sum_n G_{nl}(R) \exp\big(in[\psi + (\pi/2)]\big), \qquad (4.5.2.12)$$

where, as usual, the sum is over only those values of n that satisfy the helix selection rule. Using equations (4.5.2.8) and (4.5.2.10) shows that the Fourier–Bessel structure factors may be written in terms of the atomic coordinates as

$$G_{nl}(R) = \sum_j f_j(\rho) J_n(2\pi Rr_j) \exp\big(i[-n\varphi_j + (2\pi lz_j/c)]\big), \qquad (4.5.2.13)$$

where $f_j(\rho)$ is the (spherically symmetric) atomic scattering factor (usually including an isotropic temperature factor) of the jth

atom and $\rho = (R^2 + l^2/Z^2)^{1/2}$ is the spherical radius in reciprocal space. Equations (4.5.2.12) and (4.5.2.13) allow the complex diffracted amplitudes for a helical molecule to be calculated from the atomic coordinates, and are analogous to expressions for the structure factors in conventional crystallography.

The significance of the selection rule is now more apparent. On a particular layer plane l, not all Fourier–Bessel structure factors $G_{nl}(R)$ contribute; only those whose Bessel order n satisfies the selection rule for that value of l contribute. Since any molecule has a maximum radius, denoted here by $r_{\max}$, and since $J_n(x)$ is small for $x < |n| - 2$ and diffraction data are measured out to only a finite value of R, reference to equation (4.5.2.10) [or equation (4.5.2.13)] shows that there is a maximum Bessel order that contributes significant value to equation (4.5.2.12) (Crowther *et al.*, 1970; Makowski, 1982), so that the infinite sum over n in equation (4.5.2.12) can be replaced by a finite sum. On each layer plane there is also a minimum value of $|n|$, denoted by $n_{\min}$, that satisfies the helix selection rule, so that the region $R < R_{\min}$ is devoid of diffracted amplitude where

$$R_{\min} = \frac{n_{\min} - 2}{2\pi r_{\max}}. \qquad (4.5.2.14)$$

The selection rule therefore results in a region around the Z axis of reciprocal space that is devoid of diffraction, the shape of the region depending on the helix symmetry.

4.5.2.3.3. Approximate helix symmetry

In some cases the nature of the subunits and their interactions results in a structure that is not exactly periodic. Consider a helical structure with $u + x$ subunits in v turns, where x is a small ($x \ll 1$) real number; *i.e.* the structure has approximate, but not exact, u_v helix symmetry. Since the molecule has an *approximate* repeat distance c, only those layer planes close to those at $Z = l/c$ show significant diffraction. Denoting by Z_{mn} the Z coordinate of the nth Bessel order and its associated value of m, and using the selection rule shows that

$$Z_{mn} = [(um + vn)/c] + (mx/c) = (l/c) + (mx/c), \qquad (4.5.2.15)$$

so that the positions of the Bessel orders are shifted by mx/c from their positions if the helix symmetry is exactly u_v. At moderate resolution m is small so the shift is small. Hence Bessel orders that would have been coincident on a particular layer plane are now separated in reciprocal space. This is referred to as *layer-plane splitting* and was first observed in fibre diffraction patterns from tobacco mosaic virus (TMV) (Franklin & Klug, 1955). Splitting can be used to advantage in structure determination (Section 4.5.2.6.6).

As an example, TMV has approximately 49_3 helix symmetry with a c repeat of 69 Å. However, close inspection of diffraction patterns from TMV shows that there are actually about 49.02 subunits in three turns (Stubbs & Makowski, 1982). The virus is therefore more accurately described as a 2451_{150} helix with a c repeat of 3450 Å. The layer lines corresponding to this larger repeat distance are not observed, but the effects of layer-plane splitting are detectable (Stubbs & Makowski, 1982).

4.5.2.4. Diffraction by fibres

The kind of diffraction pattern obtained from a fibre specimen made up of helical molecules depends on the kind of specimen as described in Section 4.5.2.2. This section is divided into four parts. The first two describe diffraction patterns obtained from noncrystalline and polycrystalline fibres (which are the most common kinds used for structural analysis), and the last two describe diffraction by partially crystalline fibres.

4.5.2.4.1. Noncrystalline fibres

A noncrystalline fibre is made up of a collection of helical molecules that are oriented parallel to each other, but are otherwise randomly positioned and rotated relative to each other. The recorded intensity, $I_l(R)$, is therefore that diffracted by a single molecule cylindrically averaged about the Z axis in reciprocal space *i.e.*

$$I_l(R) = (1/2\pi)\int_0^{2\pi} |F_l(R, \psi)|^2 \, \mathrm{d}\psi; \qquad (4.5.2.16)$$

using equation (4.5.2.12) shows that

$$I_l(R) = \sum_n |G_{nl}(R)|^2, \qquad (4.5.2.17)$$

where, as usual, the sum is over the values of n that satisfy the helix selection rule. On the diffraction pattern, reciprocal space (R, ψ, Z) collapses to the two dimensions (R, Z). The R axis is called the *equator* and the Z axis the *meridian*. The layer planes collapse to *layer lines*, at $Z = l/c$, which are indexed by l. Equation (4.5.2.17) gives a rather simple relationship between the recorded intensity and the Fourier–Bessel structure factors.

Coherence length and disorientation, as described in Section 4.5.2.2, also affect the form of the diffraction pattern. These effects are described here, although they also apply to other than noncrystalline fibres. A finite coherence length leads to smearing of the layer lines along the Z direction. If the average coherence length of the molecules is l_c, the intensity distribution $I_l(R, Z)$ about the lth layer line can be approximated by

$$I_l(R, Z) = I_l(R) \exp\left(-\pi l_c^2 [Z - (l/c)]^2\right). \qquad (4.5.2.18)$$

It is convenient to express the effects of disorientation on the intensity distribution of a fibre diffraction pattern by writing the latter as a function of the polar coordinates (ρ, σ) (where σ is the angle with the Z axis) in (R, Z) space. Assuming a Gaussian orientation density function [equation (4.5.2.1)], if α_0 is small and the effects of disorientation dominate over those of coherence length (which is usually the case except close to the meridian), then the distribution of intensity about one layer line can be approximated by (Holmes & Barrington Leigh, 1974; Stubbs, 1974)

$$I(\rho, \sigma) \simeq \frac{I_l(R)}{2\pi\alpha_0 l_c \rho} \exp\left[-\frac{(\sigma - \sigma_l)^2}{2\beta^2}\right], \qquad (4.5.2.19)$$

where (Millane & Arnott, 1986; Millane, 1989*c*)

$$\beta^2 = \alpha_0^2 + (1/2\pi l_c^2 \rho^2 \sin^2 \sigma_l) \qquad (4.5.2.20)$$

and σ_l is the polar angle at the centre of the layer line, *i.e.* $R = \rho \sin \sigma_l$. The effect of disorientation, therefore, is to smear each layer line about the origin of reciprocal space.

4.5.2.4.2. Polycrystalline fibres

A polycrystalline fibre is made up of crystallites that are oriented parallel to each other, but are randomly positioned and randomly rotated about their molecular axes. The recorded diffraction pattern is the intensity diffracted by a single crystallite, cylindrically averaged about the Z axis. On a fibre diffraction pattern, therefore, the Bragg reflections are cylindrically projected onto the (R, Z) plane and their positions are described

by the cylindrically projected reciprocal lattice (Finkenstadt & Millane, 1998).

The molecules are periodic and are therefore usually aligned with one of the unit-cell vectors. Since the z axis is defined as the fibre axis, it is usual in fibre diffraction to take the **c** lattice vector as the unique axis and as the lattice vector parallel to the molecular axes. It is almost always the case that the fibre is rotationally disordered about the molecular axes, *i.e.* about **c**. Consider first the case of a monoclinic unit cell ($\alpha = \beta = 90^\circ$) so that the reciprocal lattice is cylindrically projected about $\mathbf{c}^*$. The cylindrical coordinates of the projected reciprocal-lattice points are then given by

$$R_{hkl}^2 = h^2a^{*2} + k^2b^{*2} + 2hka^*b^*\cos\gamma^* \qquad (4.5.2.21)$$

and

$$Z_{hkl} = lc^*, \qquad (4.5.2.22)$$

so that R depends only on h and k, and Z depends only on l. Reflections with fixed h and k lie on straight *row lines*. Certain sets of distinct reciprocal-lattice points will have the same value of R_{hkl} and therefore superimpose in cylindrical projection. For example, for an orthorhombic system ($\gamma = 90^\circ$) the reciprocal-lattice points (hkl), $(\bar{h}kl)$, $(h\bar{k}l)$ and $(\bar{h}\bar{k}l)$ superimpose. Furthermore, the crystallites in a fibre specimen are usually oriented randomly up and down so that the reciprocal-lattice points (hkl) and $(hk\bar{l})$ superimpose, so that in the orthorhombic case eight reciprocal-lattice points superimpose. Also, as described below, reciprocal-lattice points that have similar values of R can effectively superimpose.

If the unit cell is either triclinic, or is monoclinic with $\alpha \neq 90^\circ$ or $\beta \neq 90^\circ$, then $\mathbf{c}^*$ is inclined to **c** and the Z axis, and the reciprocal lattice is not cylindrically projected about $\mathbf{c}^*$. Equation (4.5.2.22) for Z_{hkl} still applies, but the cylindrical radius is given by

$$R_{hkl}^2 = h^2a^{*2} + k^2b^{*2} + l^2[c^{*2} - (1/c^2)] + 2hka^*b^*\cos\gamma^* + 2hla^*c^*\cos\beta^* + 2klb^*c^*\cos\alpha^* \qquad (4.5.2.23)$$

and the row lines are curved (Finkenstadt & Millane, 1998).

The most complicated situation arises if the crystallites are rotationally disordered about an axis that is inclined to **c**. Reciprocal space is then rotated about an axis that is inclined to the normal to the a^*b^* plane, R_{hkl} and Z_{hkl} are both functions of h, k and l, equation (4.5.2.23) does not apply, and reciprocal-lattice points for fixed l do not lie on layer lines of constant Z. Although this situation is rather unusual, it does occur (Daubeny *et al.*, 1954), and is described in detail by Finkenstadt & Millane (1998).

The observed fibre diffraction pattern consists of reflections at the projected reciprocal-lattice points whose intensities are equal to the sums of the intensities of the contributing structure factors. The observed intensity, denoted by $I_l(R_{hk})$, at a projected reciprocal-lattice point on the lth layer line and with $R = R_{hk}$ is therefore given by (assuming, for simplicity, a monoclinic system)

$$I_l(R_{hk}) = \sum_{h',\,k' \in \mathscr{S}(h,\,k)} |F_{h'k'l}|^2, \qquad (4.5.2.24)$$

where $\mathscr{S}(h, k)$ denotes the set of indices (h', k') such that $R_{h'k'} = R_{hk}$. The number of independent reflections contributing in equation (4.5.2.24) depends on the space-group symmetry of the crystallites, because of either systematic absences or structure factors whose values are related.

The effect of a finite crystallite size is to smear what would otherwise be infinitely sharp reflections into broadened reflections of a finite size. If the average crystallite dimensions normal and parallel to the z axis are $l_{\rm lat}$ (*i.e.* in the 'lateral' direction) and $l_{\rm axial}$ (*i.e.* in the 'axial' direction), respectively, the profile of the reflection centred at $(R_{hk}, Z = l/c)$ can be written as (Fraser *et al.*, 1984; Millane & Arnott, 1986; Millane, 1989*c*)

$$I(R, Z) = I_l(R_{hk})S(R - R_{hk}, Z - l/c), \qquad (4.5.2.25)$$

where the profile function $S(R, Z)$ can be approximated by

$$S(R, Z) = \exp[-\pi(l_{\rm lat}^2R^2 + l_{\rm axial}^2Z^2)]. \qquad (4.5.2.26)$$

The effect of crystallite disorientation is to smear the reflections given by equation (4.5.2.26) about the origin of the projected reciprocal space. If the effects of disorientation dominate over those of crystallite size, then the profile of a reflection can be approximated by (Fraser *et al.*, 1984; Millane & Arnott, 1986; Millane, 1989*c*)

$$I(\rho, \sigma) \simeq \frac{I_l(R_{hk})}{2\pi\alpha_0 l_{\rm lat} l_{\rm axial}\rho} \times \exp\left(-\left[\frac{(\rho - \rho_{hkl})^2}{2\beta_\rho^2} + \frac{(\sigma - \sigma_{hkl})^2}{2\beta_\sigma^2}\right]\right), \qquad (4.5.2.27)$$

where $(\rho_{hkl}, \sigma_{hkl})$ are the polar coordinates of the reflection,

$$\beta_\rho^2 = \frac{l_{\rm lat}^2 l_{\rm axial}^2}{2\pi(l_{\rm lat}^2\sin^2\sigma_{hkl} + l_{\rm axial}^2\cos^2\sigma_{hkl})} \qquad (4.5.2.28)$$

and

$$\beta_\sigma^2 = \alpha_0 + \frac{l_{\rm lat}^2 l_{\rm axial}^2}{2\pi\rho_{hkl}^2(l_{\rm axial}^2\sin^2\sigma_{hkl} + l_{\rm lat}^2\cos^2\sigma_{hkl})}. \qquad (4.5.2.29)$$

Reflections that have similar enough (R, Z) coordinates overlap severely with each other and are also included in the sum in equation (4.5.2.24). This is quite common in practice because a number of sets of reflections may have similar values of R_{hk}.

4.5.2.4.3. Random copolymers

Random copolymers are made up of a small number of different *kinds* of monomer, whose sequence along the polymer chain is not regular, but is random, or partially random. A particularly interesting class are synthetic polymers such as copolyesters that form a variety of liquid-crystalline phases and have useful mechanical properties (Biswas & Blackwell, 1988*a*). The structures of these materials have been studied quite extensively using X-ray fibre diffraction analysis. Because the molecules do not have an average c repeat, their diffraction patterns do not consist of equally spaced layer lines. However, as a result of the small number of distinct spacings associated with the monomers, diffracted intensity is concentrated about layer lines, but these are irregularly spaced (along Z) and are *aperiodic*. Since the molecule is not periodic, the basic theory of diffraction by helical molecules described in Section 4.5.2.3.2 does not apply in this case. Cylindrically averaged diffraction from random copolymers is described here. Related approaches have been described independently by Hendricks & Teller (1942) and Blackwell *et al.* (1984). Hendricks & Teller (1942) considered the rather general problem of diffraction by layered structures made up of different kinds of layers, the probability of a layer at a

particular level depending on the layers present at adjacent levels. This is a one-dimensional disordered structure that can be used to describe a random copolymer. Blackwell and co-workers have developed a similar theory in terms of a one-dimensional paracrystalline model (Hosemann & Bagchi, 1962) for diffraction by random copolymers (Blackwell *et al.*, 1984; Biswas & Blackwell, 1988*a*), and this is the theory described here.

Consider a random copolymer made up of monomer units (residues) of N different types. Since the disorder is along the length of the polymer, some of the main characteristics of diffraction from such a molecule can be elucidated by studying the diffraction along the meridian of the diffraction pattern. The meridional diffraction is the intensity of the Fourier transform of the polymer chain projected onto the z axis and averaged over all possible monomer sequences. The diffraction pattern depends on the monomer (molar) compositions, denoted by p_i, the statistics of the monomer sequence (described by the probability of the different possible monomer pairs in this model) and the Fourier transform of the monomer units. Development of this model shows that the meridional diffracted intensity $I(Z)$ can be written in the form (Blackwell *et al.*, 1984; Biswas & Blackwell, 1988*a*; Schneider *et al.*, 1991)

$$I(Z) = \sum_i p_i |F_i(Z)|^2 + 2\sum_i \sum_{j\neq i} \Re\{F_i(Z) F_j^*(Z) T_{ij}(Z)\}, \quad (4.5.2.30)$$

where the summations are over the different monomer types and $F_i(Z)$ is the axial Fourier transform of the ith monomer unit (each referenced to a common origin). The $T_{ij}(Z)$ are most conveniently described by defining them as the ijth element of an $N \times N$ matrix $\mathbf{T}$, which is given by

$$\mathbf{T} = \frac{\mathbf{PMX}(Z)}{\mathbf{I} - \mathbf{MX}(Z)}, \quad (4.5.2.31)$$

where $\mathbf{P}$, $\mathbf{M}$, $\mathbf{X}(Z)$ and $\mathbf{I}$ are $N \times N$ matrices. $\mathbf{I}$ is the unit matrix and $\mathbf{P}$ is a diagonal matrix with elements p_i. The elements M_{ij} of $\mathbf{M}$ are the probabilities of forming ij monomer pairs and can be generated for different kinds of random sequence (*e.g.* chemical restrictions on the occurrence of particular monomer pairs, random chains, varying degrees of blockiness, tendency towards alternating sequences *etc.*) (Schneider *et al.*, 1991). The matrix $\mathbf{X}(Z)$ is diagonal with elements equal to $\langle \exp(i2\pi Z\zeta_i) \rangle$ where the ζ_i are the projected monomer lengths and the average is over all chain conformations.

Equation (4.5.2.30) can be used to calculate the meridional diffraction for a particular random copolymer. The most important result of such a calculation is that intensity maxima are spaced irregularly along the meridian. The positions of the maxima depend on the monomer proportions, the sequence statistics and the projected monomer lengths.

The full cylindrically averaged diffraction pattern, denoted by $I(R, Z)$, from a noncrystalline specimen containing oriented random copolymer chains can be calculated by replacing $F_i(Z)$ in equation (4.5.2.30) by $F_i(R, Z)$, *i.e.*

$$I(R, Z) = \sum_i p_i |F_i(R, Z)|^2 + 2\sum_i \sum_{j\neq i} \Re\{F_i(R, Z) F_j^*(R, Z) T_{ij}(Z)\} \quad (4.5.2.32)$$

(Biswas & Blackwell, 1988*a*). Note that we write $I(R, Z)$ rather than $I_l(Z)$ since the pattern cannot be indexed on the basis of regularly spaced layer lines. The $F_i(R, Z)$ in equation (4.5.2.32) depend on the chain conformation, since this affects the range of monomer orientations and hence their average diffraction. Chivers & Blackwell (1985) have considered two extreme cases, one corresponding to fixed conformations between monomers and the other corresponding to completely random conformations between monomers, and have derived expressions for the diffracted intensity in both cases. Equation (4.5.2.32) allows one to calculate the fibre diffraction pattern from an array of parallel random copolymers that exhibit no lateral ordering. The diffraction pattern consists of irregularly spaced layer lines whose spacings (in Z) are the same as those described above for the meridional maxima. Measurement of layer-line spacings and intensities and comparison with calculations based on the constituent monomers allows chain conformations to be estimated (Biswas & Blackwell, 1988*a*).

Diffraction patterns from liquid-crystalline random copolymers often contain sharp Bragg maxima on the layer lines. This indicates that, despite the random sequence and the possible dissimilarity of the component monomers, the chains are able to pack together in a regular way (Biswas & Blackwell, 1988*b*,*c*). Expressions that allow calculation of diffraction patterns for arrays of polymers with minimal registration, in which short, non-identical sequences form layers, have been derived (Biswas & Blackwell, 1988*b*,*c*). Calculation of diffracted intensities, coupled with molecular-mechanics modelling, allows chain conformations and packing to be investigated (Hofmann *et al.*, 1994).

4.5.2.4.4. Partially crystalline fibres

In this section we address disorder in the packing of the molecules in a polycrystalline fibre. The presence of disorder within the crystallites modifies the intensities of the Bragg reflections, as well as introducing continuous diffraction. The dominant effect, sometimes seen on fibre diffraction patterns (Stroud & Millane, 1995*a*), is for Bragg reflections to remain at low resolution but to be replaced by continuous diffraction at high resolution. There are two distinct cases to consider. The first is where the distortions at different lattice points in the crystallite are uncorrelated, and the second is where they are correlated.

Disorder within a crystallite in a polycrystalline fibre may consist of (1) deviations in the positions of the molecules (which are treated as rigid bodies) from their positions in a regular lattice, (2) rotations of the molecules about their molecular axes from their rotational positions in an ordered crystal, and (3) random orientations (up or down) of the molecules. The first of these is called *lattice disorder*, and the second and third are components of *substitution disorder*.

Uncorrelated disorder has been treated by a number of authors (Clark & Muus, 1962; Tanaka & Naya, 1969; Fraser & MacRae, 1973; Arnott, 1980). A rather complete model has been described by Millane & Stroud (1991) and Stroud & Millane (1995*b*), which is presented here. If the lattice and substitution disorders are independent, and the lattice and substitution distortions at different lattice sites are uncorrelated, then the cylindrically averaged layer-line intensities $I_l(R)$ diffracted from a fibre can be written as a sum of Bragg and diffuse (continuous) intensities (Tanaka & Naya, 1969):

$$I_l(R) = I_l^{\mathrm{B}}(R) + I_l^{\mathrm{D}}(R). \quad (4.5.2.33)$$

The profiles of the Bragg reflections are independent of the position of the reflection in reciprocal space. If the Cartesian components of the lattice distortions are independent, normally distributed, and the x and y components have equal variances, cylindrical averaging of the diffracted intensity can be performed analytically.

The lattice disorder consists of distortions of the two-dimensional lattice (in the lateral plane) into three-dimensional space, and in the *absence* of substitution disorder the Bragg component is given by (Stroud & Millane, 1995*b*)

$$I_l^{\rm B}(R) = I_l(R_{hk}) w_{\rm lattice}(R_{hk}, l/c) \qquad (4.5.2.34)$$

where $I_l(R_{hk})$ is given by equation (4.5.2.24), the *lattice disorder weight*, $w_{\rm lattice}(R, Z)$, is given by

$$w_{\rm lattice}(R, Z) = \exp[-4\pi^2(R^2\sigma_{\rm lat}^2 + Z^2\sigma_{\rm axial}^2)], \qquad (4.5.2.35)$$

where $\sigma_{\rm lat}^2$ and $\sigma_{\rm axial}^2$ are the variances of the lattice distortions normal ('lateral') and parallel ('axial') to the z axis, respectively. The diffuse component is given by

$$I_l^{\rm D}(R) = (1/A_{\rm cryst}) I_l(R)[1 - w_{\rm lattice}(R, l/c)], \qquad (4.5.2.36)$$

where $I_l(R)$ is given by equation (4.5.2.17) and $A_{\rm cryst}$ is the average cross-sectional area of the crystallites. Inspection of equations (4.5.2.34) and (4.5.2.36) shows that the effect of the lattice disorder is to weight the amplitudes of the Bragg reflections down with increasing R and l, and to introduce a continuous intensity component whose amplitude increases with R and Z. Furthermore, the amplitude of the diffuse component relative to the Bragg component is inversely proportional to $A_{\rm cryst}$, and therefore is not significant unless the crystallites are small.

If substitution disorder is also present, then equation (4.5.2.33) still applies but equations (4.5.2.34) and (4.5.2.36) are replaced by

$$I_l^{\rm B}(R) = \left|\sum_n w_{nl} G_{nl}^2(R_{hk}) \exp[in(\psi_{hk} + \pi/2)]\right|^2 w_{\rm lattice}(R_{hk}, l/c) \qquad (4.5.2.37)$$

and

$$I_l^{\rm D}(R) = (1/A_{\rm cryst}) \sum_n |G_{nl}(R)|^2 [1 - |w_{nl}|^2 w_{\rm lattice}(R, l/c)], \qquad (4.5.2.38)$$

respectively, where the *substitution disorder weight*, w_{nl}, is given by

$$w_{nl} = \int_0^{c/u} \int_0^{2\pi} p(\varphi, z) \exp[i(2\pi z l/c - n\varphi)] \, {\rm d}\varphi \, {\rm d}z. \qquad (4.5.2.39)$$

In equation (4.5.2.39), $p(\varphi, z)$ is the probability density function (p.d.f.) that describes the substitution disorder, *i.e.* the p.d.f. for a molecule being rotated by φ about and translated by z along the molecular axis relative to its position in the undistorted lattice. Inspection of equations (4.5.2.37) and (4.5.2.38) shows that the substitution disorder weights the different contributing Bessel terms differently. This can lead to quite complicated effects on the diffraction pattern for various kinds of substitution disorder, resulting in different distributions and amplitudes of Bragg and diffuse diffraction over the diffraction pattern (Stroud & Millane, 1995*b*). If one assumes either uniform or normal distributions for φ and z, then expressions can be obtained for the w_{nl} in terms of the variances of the distributions of φ and z (Stroud & Millane, 1995*b*). The cases where distortions in φ are correlated with distortions in z (*e.g.* 'screw disorder'), and directional (up and down) disorder, can also be accommodated. This model has been shown to be capable of predicting diffraction patterns which are in good agreement with those measured from some disordered polycrystalline fibres (Stroud & Millane, 1995*a*).

We consider now the case of correlated packing disorder. As a result of intermolecular contacts within a polycrystalline specimen, it is possible that distortions at one lattice site will affect the degree of distortion at neighbouring sites. Coupling between distortions at different lattice sites can be included in the model of disorder by allowing the distortions at different lattice sites to be correlated. The effect of correlated distortions on diffraction patterns is that the diffracted intensity does not separate into Bragg and diffuse components as it does in the case of uncorrelated distortions [equation (4.5.2.33)]. The intensity can be described as being diffuse on the whole diffraction pattern, with (often broad) maxima occurring at some of the reciprocal-lattice points, but with no significant maxima at other reciprocal-lattice points. The widths of the profiles of the maxima generally increase with increasing resolution, whereas the widths of the Bragg maxima resulting from uncorrelated disorder as described above are independent of resolution. A broadening of diffraction maxima with increasing resolution and blending into continuous diffraction is sometimes seen on diffraction patterns from polycrystalline fibres, indicating the presence of correlated disorder (Stroud & Millane, 1996*b*).

Correlated lattice disorder consists of correlated distortions of the two-dimensional lattice into three-dimensional space. A flexible model of crystalline disorder is that based on the perturbed lattice approach (Welberry *et al.*, 1980). While formulating perturbed lattices with only nearest-neighbour interactions is complicated, a more tractable approach is to base the statistics on an imposed correlation field (de Graaf, 1989; Stroud & Millane, 1996*a*). This approach has been used to describe cylindrically averaged diffraction from polycrystalline fibres that contain correlated lattice disorder and uncorrelated substitution disorder (Stroud & Millane, 1996*a*,*b*), and is presented here.

To develop a flexible and tractable theory for diffraction from crystallites with correlated disorder, it is necessary to formulate the problem in real space. The size and shape of a crystallite in the xy (lateral) plane is described by a *shape function* $s_{\rm lat}(\mathbf{r})$, where $\mathbf{r}$ denotes the position vector in real space, which is equal to unity inside the crystallite and zero outside. The autocorrelation of the shape function, $t(\mathbf{r})$, is given by

$$t(\mathbf{r}) = \int s(\mathbf{r}') s(\mathbf{r} + \mathbf{r}') \, {\rm d}\mathbf{r}'. \qquad (4.5.2.40)$$

The correlations between the x components, and between the y components, of the distortions at any two lattice sites are taken to be identical. The correlations between distortion vectors are defined in terms of lateral, $\rho_{\rm lat}(\mathbf{r})$, and axial, $\rho_{\rm axial}(\mathbf{r})$, correlation fields such that the correlation coefficients between components of the distortions in the x (or y) and z directions, respectively, are equal to the correlation field evaluated for $\mathbf{r}$ equal to the inter-site vector. Various functional forms for the correlation fields are possible, but exponential correlation functions are usually used (Stroud & Millane, 1996*a*). If $t(\mathbf{r})$ and the correlation fields are circularly symmetric, then cylindrical averaging of the diffracted intensity can be performed analytically.

For a polycrystalline fibre with correlated lattice disorder and uncorrelated substitution disorder, the diffracted intensity is given by (Stroud & Millane, 1996*b*)

$$\begin{aligned} I_l(R) = {} & \sum_{j,k} t(r_{jk}) w_{\rm lat}(R, r_{jk}) w_{\rm axial}(l/c, r_{jk}) \\ & \times \sum_{m,n} J_{n-m}(2\pi R r_{jk}) \Re\{w_{ml} w_{nl}^* G_{ml}(R) G_{nl}^*(R) \\ & \times \exp[i(m-n)\varphi_{jk}]\} \end{aligned} \qquad (4.5.2.41)$$

where $r = |\mathbf{r}|$, the sum over (j, k) is over all the sites of the undistorted lattice within the region occupied by the autocorrelation function, (r_{jk}, φ_{jk}) are the polar coordinates of the

lattice sites, and the lateral and axial lattice disorder weights are given by

$$w_{\rm lat}(R, r) = \exp\left(-4\pi^2 R^2 \sigma_{\rm lat}^2 [1 - \rho_{\rm lat}(r)]\right) \qquad (4.5.2.42)$$

and

$$w_{\rm axial}(Z, r) = \exp\left(-4\pi^2 Z^2 \sigma_{\rm axial}^2 [1 - \rho_{\rm axial}(r)]\right). \qquad (4.5.2.43)$$

Equation (4.5.2.41) is an expression for the continuous intensity distribution along the layer lines and does not separate into Bragg and continuous components as in the case of uncorrelated disorder. However, calculations using these expressions show that the continuous intensity is sharply peaked around the projected reciprocal-lattice points at low resolution, the peaks broadening with increasing resolution until they have the character of continuous diffraction at high resolution (Stroud & Millane, 1996*a*). This is consistent with the character of diffraction patterns from some disordered polycrystalline fibres. A detailed study of the effects of correlated disorder on fibre diffraction patterns, and analysis of such disorder, can be found in Stroud & Millane (1996*a*) and Stroud & Millane (1996*b*).

4.5.2.5. Processing diffraction data

Since the diffraction pattern from a fibre is two-dimensional, it can be collected with a single exposure of a stationary specimen. Diffraction data are collected either on film, which is subsequently scanned by a two-dimensional microdensitometer to obtain a digitized representation of the diffracted intensity, or using an electronic area detector (imaging plate, CCD camera, wire detector *etc.*) (Fraser *et al.*, 1976; Namba, Yamashita & Vonderviszt, 1989; Lorenz & Holmes, 1993). We assume here that the diffraction pattern is recorded on a flat film (or detector) that is normal to the incident X-ray beam, although other film geometries are easily accommodated (Fraser *et al.*, 1976). The fibre specimen is usually oriented with its axis normal to the incident X-ray beam, although, as is described below, it is sometimes tilted by a small angle to the normal in order to better access reciprocal space close to the meridian. The diffraction and camera geometry are shown in Fig. 4.5.2.1. Referring to this figure, P and S denote the intersections of the diffracted beam with the sphere of reflection and the film, respectively. The fibre, and therefore reciprocal space, is tilted by an angle β to the normal to the incident beam. The angles μ and χ define the direction of the diffracted beam and θ is the Bragg angle. Cartesian and polar coordinates on the film are denoted by (u, v) and (r, φ), respectively, and D denotes the film-to-specimen distance.

Inspection of Fig. 4.5.2.1 shows that the cylindrical (R, ψ, Z) and spherical (ρ, ψ, σ) polar coordinates in reciprocal space are related to μ and χ by

$$\rho = (1/\lambda)[2(1 - \cos\mu - \cos\chi)]^{1/2}, \qquad (4.5.2.44)$$

$$Z = (1/\lambda)[\sin\beta(1 - \cos\mu\cos\chi) + \cos\beta\sin\chi]^{1/2}, \qquad (4.5.2.45)$$

$$R = (\rho^2 - Z^2)^{1/2}, \qquad (4.5.2.46)$$

$$\sin\psi = \frac{\sin\mu\cos\chi}{R\lambda} \qquad (4.5.2.47)$$

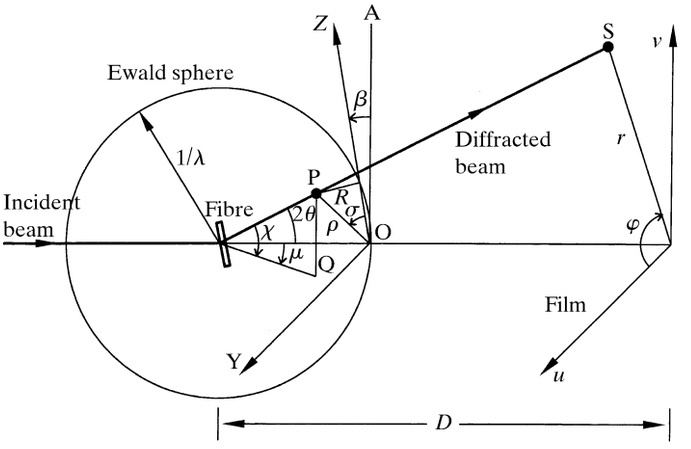


Fig. 4.5.2.1. Fibre diffraction geometry (see text). O is the origin of reciprocal space and **OA** is normal to the incident X-ray beam. Reciprocal space is rotated about **OY** so that the Z axis is inclined at an angle β to **OA**. Q is the projection of P onto the plane containing the incident beam and **OY**.

and

$$\tan\sigma = R/Z. \qquad (4.5.2.48)$$

The coordinates on the film are related to μ and χ by

$$u = D\tan\mu \qquad (4.5.2.49)$$

and

$$v = D\cos\mu\tan\chi, \qquad (4.5.2.50)$$

and we also have that

$$r = D\tan 2\theta. \qquad (4.5.2.51)$$

Use of the above equations allows the reciprocal-space coordinates to be calculated from film-space coordinates, and *vice versa.* The film coordinates (u, v) represent a relatively undistorted map of reciprocal space (R, Z), except near the v (vertical) axis of the diffraction pattern. The meridian of reciprocal space does not map onto the film. Inspection of Fig. 4.5.2.1 shows that the only point on the meridian that does appear on the film is at $Z = \lambda^{-1}\sin\beta$. The region *close* to the meridian that appears on the film can therefore be manipulated by adjusting the fibre tilt.

The film-to-specimen distance can be determined by including with the specimen a crystalline power that gives a diffraction ring of known spacing and adjusting the film-to-specimen distance so that the calculated and observed rings coincide. A nonzero fibre tilt leads to differences between the upper and lower halves of the diffraction pattern, and these differences can be used to determine the tilt. This can be done by either calculating the ρ and χ values for several sets of the same reflection above and below the equator and using the relationship

$$\tan\beta = \frac{\sin\chi_{\rm U} + \sin\chi_{\rm L}}{\lambda^2\rho^2}, \qquad (4.5.2.52)$$

where $\chi_{\rm U}$ and $\chi_{\rm L}$ refer to the upper and lower ($\chi < 0$) reflections (Millane & Arnott, 1986; Lorenz & Holmes, 1993), or by finding the tilt that minimizes the differences between optical densities at the same reciprocal-space coordinates above and below the

equator (Fraser *et al.*, 1976). The optical densities may also be corrected for the effects of film (or detector) nonlinearity (Fraser *et al.*, 1976) and the effects of variable absorption owing to the oblique passage of the beam through the film using expressions given by Fraser *et al.* (1976) and Lorenz & Holmes (1993).

Accurate subtraction of background diffraction is important in order to obtain accurate intensity measurements. One approach to estimating background diffraction is to fit a global background function, usually expanded as a polynomial (Lorenz & Holmes, 1993) or a Fourier–Bessel (Millane & Arnott, 1985) series, to optical densities at a set of points on the diffraction pattern that represent background alone. The background function may or may not be circularly symmetric. The background function is subtracted from the whole diffraction pattern. Another approach, suitable only for Bragg diffraction patterns, is to fit a plane under each reflection, either to the peripheral regions of the reflection or as part of a profile-fitting procedure (Fraser *et al.*, 1976). A different plane is required for each reflection. A third approach, more suitable for continuous diffraction patterns, is to fit a one-dimensional polynomial in angle, for each value of r on the film, possibly as part of a deconvolution procedure (Makowski, 1978). Recently, Ivanova & Makowski (1998) have described an iterative low-pass filtering technique for estimating the background on diffraction patterns from poorly oriented specimens in which there is little space between the layer lines for sampling the background.

A polarization correction is applied to the diffraction pattern, where for unpolarized X-rays (laboratory sources) the polarization factor p is given by (Fraser *et al.*, 1976)

$$p = (1 + \cos^2 2\theta)/2. \tag{4.5.2.53}$$

The diffraction pattern is usually mapped into reciprocal space (R, Z) for subsequent analysis. The mapping is performed by assigning to the intensity $I(R, Z)$ at position (R, Z) in reciprocal space the value given by (Fraser *et al.*, 1976)

$$I(R, Z) = \frac{I(u(R, Z); v(R, Z))}{\cos^3 \mu \cos^3 \chi}, \tag{4.5.2.54}$$

where $I(u; v)$ denotes the intensity on the film. The functions $u(R, Z)$ and $v(R, Z)$ can be derived from the equations given above. Note that equation (4.5.2.54) includes, implicitly, the Lorentz factor.

Subsequent processing depends on whether the diffraction pattern is continuous (*i.e.* from a noncrystalline specimen) or Bragg (*i.e.* from a polycrystalline specimen). Diffraction patterns from partially crystalline specimens that contain both components have been analysed using a combination of both approaches (Arnott *et al.*, 1986; Park *et al.*, 1987).

For a diffraction pattern containing continuous diffraction on layer lines, one usually extracts the cylindrically averaged transform $I_l(R)$ from the intensity $I(R, Z)$ on the diffraction pattern mapped into reciprocal space. This involves correcting for the effects of coherence length and disorientation expressed by equation (4.5.2.19), and for the overlap of the smeared layer lines that results from their increasing width with increasing R. The diffracted intensity $I(\rho, \sigma)$ in polar coordinates in reciprocal space is equal to the sum of the diffraction $I_l(\rho, \sigma)$ due to each (overlapping) layer line so that

$$I(\rho, \sigma) = \sum_l I_l(\rho, \sigma). \tag{4.5.2.55}$$

Referring to equations (4.5.2.55), (4.5.2.19) and (4.5.2.20) shows that if the smearing due to disorientation dominates over that due to coherence length, then for fixed ρ, equation (4.5.2.55) represents a convolution along σ of the layer-line intensities $I_l(\rho \sin \sigma_l)$ with the Gaussian angular profile in equation (4.5.2.19). By mapping the intensity $I(R, Z)$ into polar coordinates as $I(\rho, \sigma)$, or by simply sampling $I(R, Z)$ for fixed ρ and equally spaced samples of σ, $I_l(R)$ can be calculated from $I(\rho, \sigma)$ by deconvolution, usually by some appropriate solution of the resulting system of linear equations (Makowski, 1978). If the effects of coherence length are significant, as they often are, then equation (4.5.2.55) does not represent a convolution since the width of the Gaussian smearing function depends on σ through equation (4.5.2.20). However, the problem can still be posed as the solution of a system of linear equations and becomes one of profile fitting rather than deconvolution (Millane & Arnott, 1986). This allows the layer-line intensities to be extracted from the data beyond the resolution where they overlap, although there is a limiting resolution, owing to excessive overlap, beyond which reliable data cannot be obtained (Makowski, 1978; Millane & Arnott, 1986). This procedure requires that α_0 and l_c be known; these parameters can be estimated from the angular profiles at low resolution where there is no overlap, or they can be determined as part of the profile-fitting procedure.

For a diffraction pattern from a polycrystalline specimen containing Bragg reflections, the intensities $I_l(R_{hk})$ given by equation (4.5.2.24) need to be extracted from the intensity $I(R, Z)$ on the diffraction pattern mapped into reciprocal space. Each composite reflection $I_l(R_{hk})$ is smeared into a spot whose intensity profile is given by equation (4.5.2.27), and adjacent reflections may overlap. The intensity $I_l(R_{hk})$ is equal to the intensity $I(R, Z)$ integrated over the region of the spot, and the intensity at the centre of a spot is reduced, relative to $I_l(R_{hk})$, by a factor that increases with the degree of smearing.

The c repeat can be obtained immediately from the layer-line spacing. Initial estimates of the remaining cell constants can be made from inspection of the (R, Z) coordinates of low-order reflections. These values are refined by minimizing the difference between the calculated and measured (R, Z) coordinates of all the sharp reflections on the pattern.

One approach to measuring the intensities of Bragg reflections is to estimate the boundary of each spot (or a fixed proportion of the region occupied by each spot) and integrate the intensity over that region (Millane & Arnott, 1986; Hall *et al.*, 1987). For spots that overlap, an integration region that is the union of the region occupied by each contributing spot can be used, allowing the intensities for composite spots to be calculated (Millane & Arnott, 1986). This is more accurate than methods based on the measurement of the peak intensity followed by a correction for smearing. Integration methods suffer from problems associated with determining accurate spot boundaries and they are not capable of separating weakly overlapping spots. A more effective approach is one based on profile fitting. The intensity distribution on the diffraction pattern can be written as

$$I(R, Z) = \sum_l \sum_{h,k} I_l(R_{hk}, R, Z), \tag{4.5.2.56}$$

where $I_l(R_{hk}, R, Z)$ denotes the intensity distribution of the spot $I_l(R_{hk})$, and the sums are over all spots on the diffraction pattern. Using equation (4.5.2.27) shows that equation (4.5.2.56) can be written as

$$I(R, Z) = \sum_l \sum_{h,k} I_l(R_{hk}) S(R_{hk}; l/c; R; Z), \tag{4.5.2.57}$$

where $S(R_{hk}; l/c; R; Z)$ denotes the profile of the spot centred at $(R_{hk}, l/c)$ [which can be derived from equation (4.5.2.27)]. Given estimates of the parameters $l_{\rm lat}$, $l_{\rm axial}$ and α_0, equation (4.5.2.57) can be written as a system of linear equations that can be solved

for the intensities $I_l(R_{hk})$ from the data $I(R, Z)$ on the diffraction pattern. The parameters $l_{\rm lat}$, $l_{\rm axial}$ and α_0, as well as the cell constants and possibly other parameters, can also be refined as part of the profile-fitting procedure using nonlinear optimization.

A suite of programs for processing fibre diffraction data is distributed (and often developed) by the Collaborative Computational Project for Fibre and Polymer Diffraction (CCP13) in the UK (http://www.ccp13.ac.uk/) (Shotton *et al.*, 1998).

4.5.2.6. Structure determination

4.5.2.6.1. Overview

Structure determination in fibre diffraction is concerned with determining atomic coordinates or some other structural parameters, from the measured cylindrically averaged diffraction data. Fibre diffraction analysis suffers from the phase problem and low resolution (diffraction data rarely extend beyond 3 Å resolution), but this is no worse than in protein crystallography where phases derived from, say, isomorphous replacement or molecular replacement, coupled with the considerable stereochemical information usually available on the molecule under study, together contribute enough information to lead to precise structures. What makes structure determination by fibre diffraction more difficult is the loss of information owing to the cylindrical averaging of the diffraction data. However, in spite of these difficulties, fibre diffraction has been used to determine, with high precision, the structures of a wide variety of biological and synthetic polymers, and other macromolecular assemblies. Because of the size of the repeating unit and the resolution of the diffraction data, methods for structure determination in fibre diffraction tend to mimic those of macromolecular (protein) crystallography, rather than small-molecule crystallography (direct methods).

For a noncrystalline fibre one can determine only the molecular structure from the continuous diffraction data, whereas for a polycrystalline fibre one can determine crystal structures from the Bragg diffraction data. However, there is little fundamental difference between methods used for structure determination with noncrystalline and polycrystalline fibres. For partially crystalline fibres, little has so far been attempted with regard to rigorous structure determination.

As is the case with protein crystallography, the precise methods used for structure determination by fibre diffraction depend on the particular problem at hand. A variety of tools are available and one selects from these those that are appropriate given the data available in a particular case. For example, the structure of a polycrystalline polynucleotide might be determined by using Patterson functions to determine possible packing arrangements, molecular model building to define, refine and arbitrate between structures, difference Fourier synthesis to locate ions or solvent molecules, and finally assessment of the reliability of the structure. As a second example, to determine the structure of a helical virus, one might use isomorphous replacement to obtain phase estimates, calculate an electron-density map, fit a preliminary model and refine it using simulated annealing alternating with difference Fourier analysis, and assess the results. The various tools available, together with indications of where and how they are used, are described in the following sections.

Although a variety of techniques are used to solve structures using fibre diffraction, most of the methods do fall broadly into one of three classes that depend primarily on the size of the helical repeat unit. The first class applies to molecules whose repeating units are small, *i.e.* are represented by a relatively small number of independent parameters or degrees of freedom (after all stereochemical constraints have been incorporated). The structure can then be determined by an exhaustive exploration of the parameter space using molecular model building. The first example above would belong to this class. The second class of methods is appropriate when the size of the helical repeating unit is such that its structure is described by too many variable parameters for the parameter space to be explored *a priori*. It is then necessary to phase the fibre diffraction data and construct an electron-density map into which the molecular structure can be fitted and then refined. The second example above would belong to this class. The second class of methods therefore mimics conventional protein crystallography quite closely. The third class of problems applies when the structure is large, but there are too few diffraction data to attempt phasing and the usual determination of atomic coordinates. The solution to such problems varies from case to case and usually involves modelling and optimization of some kind.

An important parameter in structure determination by fibre diffraction is the degree of overlap (that results from the cylindrical averaging) in the data. This parameter is equal to the number of significant terms in equation (4.5.2.17) or the number of independent terms in equation (4.5.2.24), and depends on the position in reciprocal space and, for a polycrystalline fibre, the space-group symmetry. The number of degrees of freedom in a particular datum is equal to twice this number (since each structure factor generally has real and imaginary parts), and is denoted in this section by m. Determination of the $G_{nl}(R)$ from the cylindrically averaged data $I_l(R)$ therefore involves separating the $m/2$ amplitudes $|G_{nl}(R)|$ and assigning phases to each. The electron density can be calculated from the $G_{nl}(R)$ using equations (4.5.2.7) and (4.5.2.11).

4.5.2.6.2. Helix symmetry, cell constants and space-group symmetry

The first step in analysis of any fibre diffraction pattern is determination of the molecular helix symmetry u_v. Only the zero-order Bessel term contributes diffracted intensity on the meridian, and referring to equation (4.5.2.6) shows that the zero-order term occurs only on layer lines for which l is a multiple of u. Therefore, inspection of the distribution of diffraction along the meridian allows the value of u to be inferred. This procedure is usually effective, but can be difficult if u is large, because the first meridional maximum may be on a layer line that is difficult to measure. This difficulty was overcome in one case by Franklin & Holmes (1958) by noting that the second Bessel term on the equator is $n = u$, estimating $G_{00}(R)$ using data from a heavy-atom derivative (see Section 4.5.2.6.6), subtracting this from $I_0(R)$, and using the behaviour of the remaining intensity for small R to infer the order of the next Bessel term [using equation (4.5.2.14)] and thence u.

Referring to equations (4.5.2.6) and (4.5.2.14) shows that the distribution of $R_{\rm min}$ for $0 < l < u$ depends on the value of v. Therefore, inspection of the intensity distribution close to the meridian often allows v to be inferred. Note, however, that the distribution of $R_{\rm min}$ does not distinguish between the helix symmetries u_v and u_{u-v}. Any remaining ambiguities in the helix symmetry need to be resolved by steric considerations, or by detailed testing of models with the different symmetries against the available data.

For a polycrystalline system, the cell constants are determined from the (R, Z) coordinates of the spots on the diffraction pattern as described in Section 4.5.2.6.4. Space-group assignment is based on analysis of systematic absences, as in conventional crystallography. However, in some cases, because of possible overlap of systematic absences with other reflections, there may be some ambiguity in space-group assignment. However, the space group can always be limited to one of a few possibilities, and ambiguities can usually be resolved during structure determination (Section 4.5.2.6.4).

4.5.2.6.3. Patterson functions

In fibre diffraction, the conventional Patterson function cannot be calculated since the individual structure-factor intensities are not available. However, MacGillavry & Bruins (1948) showed that the *cylindrically averaged Patterson function* can be calculated from fibre diffraction data. Consider the function $\hat{Q}(r, z)$ defined by

$$\hat{Q}(r, z) = \sum_{l=0}^{\infty} \int_0^{\infty} \varepsilon_l I_l(R) J_0(2\pi Rr) \cos(2\pi lz/c) 2\pi R \, \mathrm{d}R, \quad (4.5.2.58)$$

where $\varepsilon_l = 1$ for $l = 0$ and 2 for $l > 0$, which can be calculated from the intensity distribution on a continuous fibre diffraction pattern. Using equations (4.5.2.7), (4.5.2.10), (4.5.2.17) and (4.5.2.58) shows that $\hat{Q}(r, z)$ is the cylindrical average of the Patterson function, $\hat{P}(r, \varphi, z)$, of one molecule, *i.e.*

$$\hat{Q}(r, z) = (1/2\pi) \int_0^{2\pi} \hat{P}(r, \varphi, z) \, \mathrm{d}\varphi. \quad (4.5.2.59)$$

The $\hat{}$ symbols on $\hat{P}(r, \varphi, z)$ and $\hat{Q}(r, z)$ indicate that these are Patterson functions of a single molecule, as distinct from the usual Patterson function of a crystal, which contains intermolecular interatomic vectors and is periodic with the same periodicity as the crystal. $\hat{P}(r, \varphi, z)$ is periodic only along z and is therefore, strictly, a Patterson function along z and an autocorrelation function along x and y (Millane, 1990*b*). The cylindrically averaged Patterson contains information on interatomic separations along the axial direction and in the lateral plane, but no information on orientations of the vectors in the lateral plane.

For a polycrystalline system; consider the function $Q(r, z)$ given by

$$Q(r, z) = \sum_l \sum_{h,k} R_{hk} I_l(R_{hk}) J_0(2\pi R_{hk} r) \cos(2\pi lz/c), \quad (4.5.2.60)$$

where the sums are over all the overlapped reflections $I_l(R_{hk})$ on the diffraction pattern, given by equation (4.5.2.24). It is easily shown that $Q(r, z)$ is related to the Patterson function $P(r, \varphi, z)$ by

$$Q(r, z) = (1/2\pi) \int_0^{2\pi} P(r, \varphi, z) \, \mathrm{d}\varphi, \quad (4.5.2.61)$$

where, in this case, $P(r, \varphi, z)$ is the usual Patterson function (expressed in cylindrical polar coordinates), *i.e.* it contains all intermolecular (both intra- and inter-unit cell) interatomic vectors and has the same translational symmetry as the unit cell. The cylindrically averaged Patterson function for polycrystalline fibres therefore contains the same information as it does for noncrystalline fibres (*i.e.* no angular information in the lateral plane), except that it also contains information on intermolecular separations.

Low resolution and cylindrical averaging, in addition to the usual difficulties with interpretation of Patterson functions, has resulted in the cylindrically averaged Patterson function not playing a major role in structure determination by fibre diffraction. However, information provided by the cylindrically averaged Patterson function has, in a number of instances, been a useful component in fibre diffraction analyses. A good review of the application of Patterson functions in fibre diffraction is given by Stubbs (1987). Removing data from the low-resolution part (or all) of the equator when calculating the cylindrically averaged Patterson function removes the strong vectors related to axially invariant (or cylindrically symmetric) parts of the map, and can aid interpretation (Namba *et al.*, 1980; Stubbs, 1987). It is also important when calculating cylindrically averaged Patterson functions to use data only at a resolution that is appropriate to the size and spacings of features one is looking for (Stubbs, 1987).

Cylindrically averaged Patterson functions were used in early applications of fibre diffraction analysis (Franklin & Gosling, 1953; Franklin & Klug, 1955). The intermolecular peaks that usually dominate in a cylindrically averaged Patterson function can help to define the locations of multiple molecules in the unit cell. Depending on the space-group symmetry, it is sometimes possible to calculate the complete three-dimensional Patterson function (or certain projections of it). This comes about because of the equivalence of the amplitudes of overlapping reflections in some high-symmetry space groups. The intensity of each reflection can then be determined and a full three-dimensional Patterson map calculated (Alexeev *et al.*, 1992). The only difficulty is that nonsystematic overlaps are often present, although these are usually relatively few in number and the intensity can be apportioned equally amongst them, the resulting errors usually being small relative to the level of detail present in the Patterson map. For lower space-group symmetries, it may not be possible to calculate a three-dimensional Patterson map, but it may be possible to calculate certain projections of the map. For example, if the overlapped $hk0$ reflections have the same intensities, a projection of the Patterson map down the c axis can be calculated. Since such a projection is along the polymer axes, it gives the relative positions of the molecules in the ab plane. If the combined helix and space-group symmetry is high, an estimate of the electron density can be obtained by averaging appropriate copies of the three-dimensional Patterson function (Alexeev *et al.*, 1992).

4.5.2.6.4. Molecular model building

The majority of the structures determined by X-ray fibre diffraction analysis have been determined by molecular model building (Campbell Smith & Arnott, 1978; Arnott, 1980; Millane, 1988). Most applications of molecular model building have been to polycrystalline systems, although there have been a number of applications to noncrystalline systems (Park *et al.*, 1987; Millane *et al.*, 1988). The approach is to use spacings and symmetry information derived directly from the diffraction pattern, coupled with the primary structure and stereochemical information on the molecule under study, to construct models of all *kinds* of possible molecular or crystal structure. These models are each refined (optimized) against the diffraction data, as well as stereochemical restraints, to produce the best model of each kind. The optimized models can be compared using various figures of merit, and in favourable cases one model will be sufficiently superior to the remainder for it to represent unequivocally the correct structure. The principle of this approach is that by making use of stereochemical constraints, the molecular and crystal structure have few enough degrees of freedom that the parameter space has a sufficiently small number of local minima for these to be identified and individually examined to find the global minimum. The X-ray phases are therefore not determined explicitly.

There are three steps involved in structure determination by molecular model building: (1) construction of all possible molecular and crystal structure models, (2) refinement of each model against the X-ray data and stereochemical restraints, and (3) adjudication among the refined models. The overall procedure for determining polymer structures using molecular model building is summarized by the flow chart in Fig. 4.5.2.2, and is described below.

The helix symmetry of the molecule, or one of a few helix symmetries, can be determined as described in Section 4.5.2.6.2. Different kinds of molecular model may correspond to one of a few different helix symmetries, usually corresponding to different

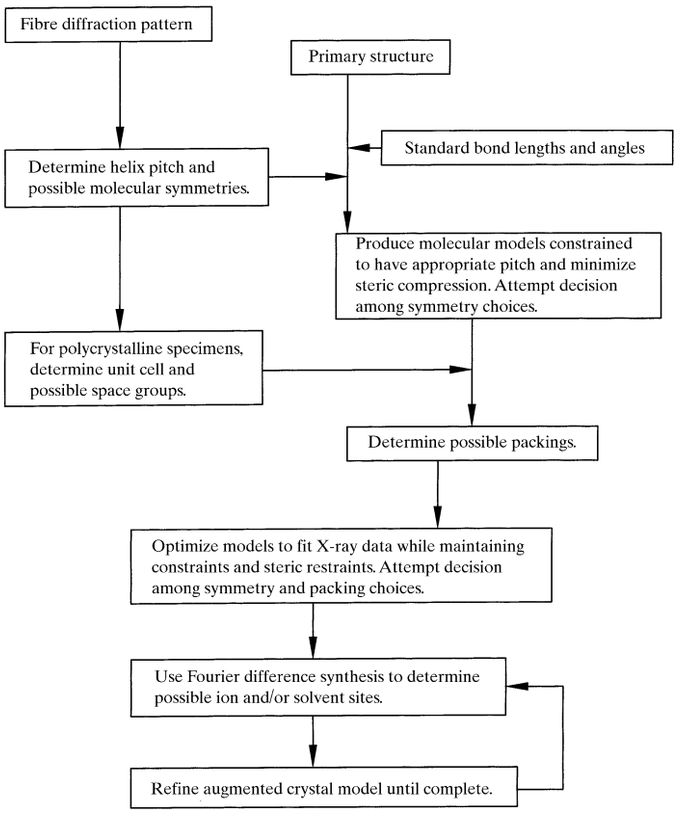

Fibre diffraction pattern
Primary structure
Standard bond lengths and angles
Determine helix pitch and possible molecular symmetries.
Produce molecular models constrained to have appropriate pitch and minimize steric compression. Attempt decision among symmetry choices.
For polycrystalline specimens, determine unit cell and possible space groups.
Determine possible packings.
Optimize models to fit X-ray data while maintaining constraints and steric restraints. Attempt decision among symmetry and packing choices.
Use Fourier difference synthesis to determine possible ion and/or solvent sites.
Refine augmented crystal model until complete.

Fig. 4.5.2.2. Flow chart of the molecular-model-building approach to structure determination (Arnott, 1980).

values of v. For example, helix symmetries u_v and u_{u-v}, which correspond to the left- and right-handed helices, cannot be distinguished on the basis of the overall intensity distribution alone. Other examples of different kinds of molecular model may include single, double or multiple helices, parallel or antiparallel double helices, different juxtapositions of chains within multiple helices and different conformational domains within the molecule. For polycrystalline systems, in addition to different kinds of molecular structures, there are often different kinds of possible packing arrangements within the unit cell. There may be a number of possible packings which correspond to different arrangements within the crystallographic asymmetric unit, and there may be more than one space group that needs to be considered.

Despite the apparent large number of potential starting models implied by the above discussion, in practice the number of feasible models is usually quite small, and many of these are often eliminated at an early stage. Definition and refinement of helical polymers [steps (1) and (2) above] are carried out using computer programs, the most popular and versatile being the *linked-atom least-squares* (*LALS*) system (Campbell Smith & Arnott, 1978; Millane *et al.*, 1985), originally developed by Arnott and co-workers in the early 1960s (Arnott & Wonacott, 1966). This system has been used to determine the structures of a wide variety of polynucleotides, polysaccharides, polyesters and polypeptides (Arnott, 1980; Arnott & Mitra, 1984; Chandrasekaran & Arnott, 1989; Millane, 1990*c*). Other refinement systems exist (Zugenmaier & Sarko, 1980; Iannelli, 1994), but the principles are essentially the same and the following discussion is in terms of the *LALS* system. The atomic coordinates are defined, using a linked-atom description, in terms of bond lengths, bond angles and conformation (torsion) angles (Campbell Smith & Arnott, 1978). Stereochemical constraints are imposed, and the number of parameters reduced, by fixing the bond lengths, often (but not always) the bond angles, and possibly some of the conformation angles. The molecular conformation is then defined by the remaining parameters. For polycrystalline systems, there are usually additional variable parameters that define the packing of the molecule(s) in the unit cell. A further source of stereochemical data is the requirement that a model exhibit no overshort nonbonded interatomic distances. These are incorporated by a quadratic nonbonded potential that is matched to a Buckingham potential (Campbell Smith & Arnott, 1978). A variety of other restraints can also be incorporated.

In the *LALS* system, the quantity Ω given by

$$\Omega = \sum_m \omega_m \Delta F_m^2 + \sum_m k_m \Delta d_m^2 + \sum_m \lambda_m G_m = X + C + L \qquad (4.5.2.62)$$

is minimized by varying a set of chosen parameters consisting of conformation angles, possibly bond angles, and packing parameters. The term X involves the differences ΔF_m between the model and experimental X-ray amplitudes – Bragg and/or continuous. The term C involves restraints to ensure that overshort nonbonded interatomic distances are driven beyond acceptable minimum values, that conformations are within desired domains, that hydrogen-bond and coordination geometries are close to the expected configurations, and a variety of other relationships are satisfied (Campbell Smith & Arnott, 1978). The ω_m and k_m are weights that are inversely proportional to the estimated variances of the data. The term L involves constraints which are relationships that are to be satisfied exactly ($G_m = 0$) and the λ_m are Lagrange multipliers. Constraints are used, for example, to ensure connectivity from one helix pitch to the next and to ensure that chemical ring systems are closed. The cost function Ω is minimized using full-matrix nonlinear least squares and singular value decomposition (Campbell Smith & Arnott, 1978).

Structure determination usually involves first using equation (4.5.2.62) with the terms C and L only, to establish the stereochemical viability of each kind of possible molecular model and packing arrangement. It is worth emphasizing that it is usually advantageous if the specimen is polycrystalline, even though the continuous diffraction contains, in principle, more information than the Bragg reflections (since the latter are sampled). This is because the molecule in a noncrystalline specimen must be refined in steric isolation, whereas for a polycrystalline specimen it is refined while packed in the crystal lattice. The extra information provided by the intermolecular contacts can often help to eliminate incorrect models. This can be particularly significant if the molecule has flexible sidechains. The initial models that survive the steric optimization are then optimized also against the X-ray data, by further refinement with X included in equation (4.5.2.62). The ratios $(\Omega_P/\Omega_Q)^{1/2}$ and $(X_P/X_Q)^{1/2}$ can be used in Hamilton's test (Hamilton, 1965) to evaluate the differences between models P and Q. On the basis of these statistical tests, one can decide if one model is superior to the others at an acceptable confidence level. In the final stages of refinement, bond angles may be varied in a 'stiffly elastic' fashion from their mean values if there are sufficient data to justify the increase in the number of degrees of freedom.

If sufficient X-ray data are available, it is sometimes possible to locate additional ordered molecules such as counterions or solvent molecules by difference Fourier synthesis as described in Section 4.5.2.6.5. Their positions can then be co-refined with the polymer structure while hydrogen bonds and coordination geometries are optimized. The resulting structure can then be used to compute improved phases to search for additional molecules. Since the signal-to-noise ratio in fibre difference syntheses is usually low, difference maps must be interpreted with caution. The assignment of counterions or solvent molecules to peaks in the difference synthesis must be supported by plausible

interactions with the rest of the structure and, following refinement of the structure, by elimination of the peak in the difference map and by a significant improvement in the agreement between the calculated and measured X-ray amplitudes.

4.5.2.6.5. Difference Fourier synthesis

Difference Fourier syntheses are widely used in both protein and small-molecule crystallography to detect structural errors or to complete partial structures (Drenth, 1994). The difficulty in applying difference Fourier techniques in fibre diffraction is that the individual observed amplitudes $|F_o|$ are not available. However, difference syntheses have found wide use in fibre diffraction analysis, one of the earliest applications being to polycrystalline fibres of polynucleotides (*e.g.* Arnott *et al.*, 1967). Calculation of a three-dimensional difference map (for the unit cell) from Bragg fibre diffraction data requires that the observed intensity $I_l(R_{hk}) = I_o$ be apportioned among the contributing intensities $|F_{hkl}|^2 = |F_o|^2$. There are two ways of doing this. The intensities may be divided equally among the contributing $(m/2)$ reflections [*i.e.* $|F_o| = (2I_o/m)^{1/2}$], or they may be divided in the same proportions as those in the model, *i.e.*

$$|F_o| = \left(\frac{I_o}{\sum|F_c|^2}\right)^{1/2} |F_c|. \qquad (4.5.2.63)$$

The advantage of the former is that it is unbiased, and the advantage of the latter is that it may be more accurate but is biased towards the model. Equal division of the intensities is often (but not always) used to minimize model bias. Once the observed amplitudes have been apportioned, an $|F_o| - |F_c|$ map can be calculated as in conventional crystallography, although noise levels will be higher owing to errors in apportioning the amplitudes. As a result of overlapping of the reflections, a synthesis based on coefficients $m|F_o| - (m-1)|F_c|$ gives a more accurate estimate of the true density than does one based on $2|F_o| - |F_c|$, as is described below. Difference syntheses for polycrystalline specimens calculated in this way have been used, for example, to locate cations and water molecules in polynucleotide and polysaccharide structures (*e.g.* Cael *et al.*, 1978), to help position molecules in the unit cell (*e.g.* Chandrasekaran *et al.*, 1994) and to help position side chains, and have also been applied in neutron fibre diffraction studies of polynucleotides (Forsyth *et al.*, 1989).

Sim (1960) has shown that the mean-squared error in difference syntheses can be minimized by weighting the coefficients based on the agreement between the calculated and observed structure amplitudes. Such an analysis has recently been conducted for fibre diffraction, and shows that the optimum difference synthesis is obtained by using coefficients (Millane & Baskaran, 1997; Baskaran & Millane, 1999*a*)

$$\left[w_m \frac{|F_c|(I_o)^{1/2}}{(\sum|F_c|^2)^{1/2}} - |F_c|\right] \exp(i\alpha_c), \qquad (4.5.2.64)$$

where m is the number of degrees of freedom as defined in Section 4.5.2.6.1. If the reflections contributing to I_o are either all centric or all acentric, then the weights are given by

$$w_m = \frac{I_{m/2}(X)}{I_{m/2-1}(X)}, \qquad (4.5.2.65)$$

where $I_m(\cdot)$ denotes the modified Bessel function of the first kind of order m, and X is given by

$$X = \frac{\kappa(I_o)^{1/2}(\sum|F_c|^2)^{1/2}}{\sum_j f_j^2}, \qquad (4.5.2.66)$$

where $\kappa = 1$ for centric reflections and 2 for acentric reflections. The form of the weighting function is more complicated if both centric and acentric reflections contribute, but it can be approximated as w' given by

$$w' = (w_{2N_a} + w_{N_c})/2, \qquad (4.5.2.67)$$

where N_a and N_c are the number of acentric and centric reflections, respectively, contributing. Use of the weighted maps reduces bias towards the model (Baskaran & Millane, 1999*b*).

For continuous diffraction data from noncrystalline specimens, the situation is essentially identical except that one works in cylindrical coordinates. Referring to equations (4.5.2.7) and (4.5.2.10), the desired difference synthesis, $\Delta g(r, \varphi, z)$, is the Fourier–Bessel transform of $G_o - G_c$ where G_o and G_c denote the observed and calculated, respectively, Fourier–Bessel structure factors $G_{nl}(R)$. Since G_o is not known, the synthesis is based on the Fourier–Bessel transform of $(|G_o| - |G_c|)\exp(i\alpha_c)$, where α_c is the phase of G_c. As in the polycrystalline case, the individual $|G_o|$ need to be estimated from the data $I_o^{1/2}$ given by equation (4.5.2.17), and can be based on either equal division of the data, or division in the same proportion as the amplitudes from the model.

Namba & Stubbs (1987*a*) have shown that the peak heights in a difference synthesis are $1/m$ times their true value, as opposed to half their true value in a conventional difference synthesis. The best estimate of the true map is therefore provided by a synthesis based on the coefficients $[m|F_o| - (m-1)|F_c|]\exp(i\alpha_c)$, rather than on $(2|F_o| - |F_c|)\exp(i\alpha_c)$. Test examples showed that the noise in the synthesis can be reduced by using a value for m that is fixed over the diffraction pattern and approximately equal to the average value of m over the pattern (Namba & Stubbs, 1987*a*). Difference Fourier maps for noncrystalline systems have been used in studies of helical viruses to locate heavy atoms, to correct errors in atomic models and to locate water molecules (Mandelkow *et al.*, 1981; Lobert *et al.*, 1987; Namba, Pattanayek & Stubbs, 1989; Wang & Stubbs, 1994).

4.5.2.6.6. Multidimensional isomorphous replacement

At low enough resolution, only one Fourier–Bessel structure factor contributes on each layer line of a fibre diffraction pattern, so that only the phase needs to be determined and the situation is no different to that in protein crystallography. If heavy-atom-derivative specimens can be prepared, the usual method of multiple isomorphous replacement (MIR) (Drenth, 1994) can be applied, which in principle requires only two heavy-atom derivatives. At higher resolution, however, more than one Fourier–Bessel structure factor contributes on each layer line. A generalized form of isomorphous replacement which involves using diffraction data from several heavy-atom derivatives to determine the real and imaginary components of each contributing $G_{nl}(R)$ is referred to as *multidimensional isomorphous replacement* (MDIR) (Namba & Stubbs, 1985). MDIR was first described and used to determine the structure of TMV at 6.7 Å resolution (Stubbs & Diamond, 1975; Holmes *et al.*, 1975), and has since been used to extend the resolution to 2.9 Å (Namba, Pattanayek & Stubbs, 1989). A consequence of cylindrical averaging is that large numbers of heavy-atom derivatives are required: at least two for each Bessel term to be separated. The theory of MDIR is outlined here.

The first step in MDIR is location of the heavy atoms in the derivative structures. The radial coordinate of a heavy atom can be determined by analysis of the intensity distribution in the low-

resolution region of the equator where only the $G_{00}(R)$ Bessel term contributes. Since $G_{00}(R)$ is real, and $I_l(R)$ can be measured continuously in R, inspection of the positions of the minima and maxima in the low-resolution region of the equator generally allows the sign of $G_{00}(R)$ to be assigned to $I_0^{1/2}(R)$, *i.e.* $G_{00}(R)$ can be determined from $I_0(R)$. If the sign is determined for both the native and a heavy-atom derivative, referring to equation (4.5.2.13) shows that

$$G_{00}^{\rm D}(R) - G_{00}(R) = o_h f_h J_0(2\pi R r_h), \tag{4.5.2.68}$$

where $G_{00}^{\rm D}(R)$ is the value derived from the derivative data, o denotes the occupancy and the subscript h denotes values for the heavy atom. The parameters o_h and r_h on the right-hand side of equation (4.5.2.68) can be searched in a trial-and-error fashion to obtain the best agreement with the left-hand side (calculated from the data) to determine the radial coordinate r_h of the heavy atom (Mandelkow & Holmes, 1974). Lobert *et al.* (1987) applied the same method to cucumber green mottle mosaic virus (CGMMV), except that the sign of $G_{00}(R)$ was taken from that of TMV.

Two approaches have been used to determine the angular and axial coordinates of the heavy atom. Mandelkow & Holmes (1974) and Holmes *et al.* (1975) used a search procedure in which the quantity $\Phi = -n\varphi_h + 2\pi l z_h/c$ is varied and used to calculate the intensity of the Fourier–Bessel structure factor for the heavy atom alone. This is compared to $I_l^{\rm D}(R) - I_l(R)$ on each layer line, where only one Bessel order contributes, and Φ chosen to minimize the mean-square difference. The values of Φ found for each layer line can then be combined to determine φ_h and z_h. In the case of CGMMV, Lobert *et al.* (1987) used the phases and Bessel-order separations from TMV to calculate Fourier–Bessel difference maps between the native and derivative data to determine the heavy-atom coordinates (r_h, φ_h, z_h).

Consider a set of isomorphous heavy-atom derivatives indexed by j. Since the analysis is applied at any point (l, R) on the fibre diffraction pattern, the symbol G_n will be used for $G_{nl}(R)$ where no confusion arises. Denote by $G_{n,j}$ the value of G_n for the jth derivative, so that

$$G_{n,j} = G_n + g_{n,j}, \tag{4.5.2.69}$$

where $g_{n,j}$ denotes the Fourier–Bessel structure factor of a structure containing the heavy atom only. Denote by A_n and B_n the real and imaginary parts, respectively, of G_n (for the native structure), and by $a_{n,j}$ and $b_{n,j}$ the real and imaginary parts of $g_{n,j}$, *i.e.* for the jth heavy-atom structure alone. Equation (4.5.2.17) can then be written as

$$I = \sum_n (A_n^2 + B_n^2) \tag{4.5.2.70}$$

for the native and

$$I_j = \sum_n [(A_n + a_{n,j})^2 + (B_n + b_{n,j})^2] \tag{4.5.2.71}$$

for the jth derivative. If intensity data are available from J heavy-atom derivatives, $a_{n,j}$ and $b_{n,j}$ can be calculated from the heavy-atom positions, and equations (4.5.2.70) and (4.5.2.71) represent a system of $J+1$ second-order equations for the m unknowns A_n and B_n. If $J+1 > m$, then the system of equations is over-determined and can be solved for the A_n and B_n. The solution of this nonlinear system can be eased by deriving a system of linear equations by substituting from (4.5.2.70) into (4.5.2.71), giving

$$\sum_n (A_n a_{n,j} + B_n b_{n,j}) = (1/2)\left[I_j - I - \sum_n (a_{n,j}^2 + b_{n,j}^2)\right]. \tag{4.5.2.72}$$

Equation (4.5.2.72) is a system of linear equations for the unknowns A_n and B_n, the solution being subject to the constraint equation (4.5.2.70). However, since the original problem is second-order, there may be up to m local minima. Stubbs & Diamond (1975) describe a numerical procedure for locating *all* the local minima and selecting the best of these based on 'continuity' of the $G_{nl}(R)$. This method was used to determine the structure of TMV at 6.7 Å resolution (Holmes *et al.*, 1975) and 4 Å resolution (Stubbs *et al.*, 1977). In current applications of MDIR a more direct solution technique is used in which the phase-determining equations (4.5.2.70) and (4.5.2.71) are solved by first solving the linear equations (4.5.2.72) by linear least squares to obtain an approximate solution, which is then refined by solving the quadratic equations (4.5.2.70) and (4.5.2.71) directly using nonlinear least squares (Namba & Stubbs, 1985).

The number of heavy-atom derivatives required can be quite demanding experimentally, although phasing with fewer heavy-atom derivatives is possible, particularly if additional information is available, such as from a related structure. The different Bessel terms may be assumed to contribute the same amplitude each, or, if the structure of a related molecule is known, the ratios of the amplitudes can be taken as being the same as those for the related molecule. Using the amplitude estimates derived using either of these two approaches, applied to both native and derivative data, the phases of the Bessel terms can be estimated using conventional MIR and data from at least two heavy-atom derivatives, allowing an initial electron-density map to be calculated. If only one heavy-atom derivative is available then two phase solutions are obtained, but the method of conventional single isomorphous replacement (SIR) (Drenth, 1994) can be used to obtain an estimate of the electron density. The electron density obtained by MIR, and particularly by SIR, in this way tends to be noisy and low contrast as a result of inaccurate division of the intensities, as well as the usual sources of errors in MIR. The electron density can, however, be improved using solvent levelling. If *no* heavy-atom derivatives are available, both the relative amplitudes *and* the phases can be based on those of a related structure. Model bias can, however, be more serious than in conventional crystallography since both the phases and the relative amplitudes are based on the model.

The feasibility of structure determination with a limited number of heavy-atom derivatives was first demonstrated by Namba & Stubbs (1987*b*) using data from TMV at 4 Å resolution. The structure of CGMMV has been determined at 5 Å resolution using data from two heavy-atom derivatives and the techniques described above (Lobert *et al.*, 1987; Lobert & Stubbs, 1990). Structure determination at this resolution using MDIR would theoretically require six heavy-atom derivatives. Initial separation of the Bessel-term amplitudes was based on the equal-amplitude assumption and also on the relative amplitudes for (homologous) TMV.

In general, the equal-amplitude assumption appears to produce reliable electron-density maps where only two or three Bessel terms contribute. The corresponding resolution depends on the helix symmetry and the molecular diameter, but can be relatively high for molecules with high helix symmetry. At higher resolution where more Bessel terms contribute, use of related or partial structures can be used to calculate initial Bessel-term amplitudes and can lead to successful phasing.

If the molecule has only approximate helix symmetry, then layer-line splitting (Section 4.5.2.3.3) can provide additional information which reduces the number of heavy-atom derivatives required. The degree of splitting is usually significantly less than

the breadth of the layer lines so that the different Bessel terms within a (split) layer line overlap. The effect of splitting can be observed, however, since the centre of a layer line, at a particular value of R, is shifted towards the position of the stronger Bessel term contributing at that radius. The shift depends on the relative magnitudes of the contributing Bessel terms, and can be measured and used in phase determination as detailed by Stubbs & Makowski (1982). If P of the heavy-atom derivatives (in addition to the native) give accurate splitting information, then an additional P linear equations [analogous to equation (4.5.2.72)] and one quadratic equation [analogous to equation (4.5.2.70)] are available for solution of the phase problem, and the number of heavy-atom derivatives required is reduced by a factor of up to two. The value of layer-line splitting was first demonstrated by recalculating an electron-density map of TMV at 6.7 Å resolution using only two derivatives, rather than using six derivatives without the use of splitting data (Stubbs & Makowski, 1982). Layer-line splitting was subsequently used in a structure determination of TMV at 3.6 Å resolution (Namba & Stubbs, 1985).

Macromolecular fibre structures that have been built into an electron-density map have been refined using both restrained least-squares (RLS) and molecular-dynamics (MD) refinements. Restrained least squares has been used to refine the structure of TMV at 2.9 Å resolution (Namba, Pattanayek & Stubbs, 1989); however, Wang & Stubbs (1993) have shown that a larger radius of convergence is obtained using MD refinement (as in protein crystallography).

Molecular-dynamics refinement in fibre diffraction has been implemented by adding a fibre diffraction option (Wang & Stubbs, 1993) to the *X-PLOR* program (Brünger, 1992). This involves including the cylindrically averaged fibre diffraction intensities in the energy term and taking account of the inter-helical subunit contacts and covalent connections in the same way as described above for RLS refinement. The effective potential-energy function E used is

$$E = E_{\rm e} + S\sum_l \sum_i w_{li}\{[I_l^{\rm o}(R_i)]^{1/2} - k[I_l^{\rm c}(R_i)]^{1/2}\}^2, \qquad (4.5.2.73)$$

where $E_{\rm e}$ is the empirical energy function (which typically includes bond-length, bond-angle and torsion-angle distortions, van der Waals and electrostatic interactions, and other terms such as ring planarity), $I_l^{\rm o}(R_i)$ and $I_l^{\rm c}(R_i)$ are the observed and calculated, respectively, cylindrically averaged diffraction intensities sampled at $R = R_i$, the w_{li} are weights for the observed intensities $I_l^{\rm o}(R_i)$ and k is a scale factor between the calculated and observed data. The quantity S is a weight to make the gradients of the two terms in equation (4.5.2.73) comparable (Wang & Stubbs, 1993), and can be estimated using the method of Brünger (1992). Molecular-dynamics refinement has been successfully used to refine the structure of CGMMV at 3.4 Å resolution (Wang & Stubbs, 1994). In the case of ribgrass mosaic virus (RMV), the close isomorphism with TMV (identical helix symmetry, similar repeat distance, significant sequence homology and similar diffraction pattern) allowed an initial model to be built based on the TMV structure, and a solution obtained at 2.9 Å by alternating molecular-dynamics refinement with difference-map and omit-map calculations (Wang *et al.*, 1997).

4.5.2.6.7. Other techniques

Aside from the techniques for structure determination described in the previous sections, a variety of other techniques have been applied to specific problems where the methods described above are not suitable. This situation usually arises where the diffraction data available are far too few, by themselves, to determine the individual atomic coordinates of a structure, even with the usual stereochemical constraints. Often only relatively low-resolution data are available, but they can be supplemented by either a low-resolution or high-resolution model of either a whole molecule or relatively large subunits. Structure determination often amounts to positioning the molecules or subunits within a larger assembly. The results can be quite precise, depending on the information available. The problem is almost always one of refinement or optimization, since it invariably involves optimizing some kind of model directly against the fibre diffraction data. The problem is usually twofold: (1) parameterizing the model with few enough parameters to obtain a usable data-to-parameter ratio, but retaining enough degrees of freedom to represent the important structural features; and (2) devising an optimization procedure that will locate the global minimum of the resulting complicated cost function. There have been numerous such applications in fibre diffraction, and rather than attempt to be exhaustive or detailed, I will briefly mention a few of the more prominent applications and techniques.

The structure of the bacteriophage Pf1 was determined at 7 Å resolution using a model in which the α-helical segments of the structure were represented by rods of electron density of appropriate dimensions and spacings (Makowski *et al.*, 1980). The positions and orientations of the rods were refined in an iterative procedure that alternated between real space and reciprocal space and also incorporated solvent levelling. Neutron fibre diffraction data have been collected from specifically deuterated phages and, starting with a model of the kind described above, iterative application of difference maps (between the deuterated and native data) was used to locate 15 (of the 46) residues, allowing construction of a model of the coat protein (Stark *et al.*, 1988; Nambudripad *et al.*, 1991).

Pf1 undergoes a temperature-induced structural transition that involves a small change in the helix symmetry. The low-temperature form has 71_{13} helix symmetry with a c repeat of 216.5 Å, and the high-temperature form (that discussed in the previous paragraph) has 27_5 helix symmetry and a c repeat of 78.3 Å. These two symmetries are very similar since $71/3 \simeq 27/5$ and $216.5/71 \simeq 78.3/27$, *i.e.* the rotations and translations from one subunit to the next are very similar in both structures.

The structure of the low-temperature form of Pf1 has been determined at 3.3 Å resolution by starting with an α-helical polyalanine model (Marvin *et al.*, 1987) and alternating rounds of molecular-dynamics refinement and model rebuilding based on $(2F_{\rm o} - F_{\rm c})$ maps and omit maps (Gonzalez *et al.*, 1995). The structure of the high-temperature form of Pf1 was determined using data to 3 Å resolution, starting with a model based on the low-temperature form, making small adjustments to satisfy the slightly different helix symmetry, and refining the model using molecular dynamics (Welsh *et al.*, 2000).

The bacteriophage Pf3 is related to Pf1 but does not undergo a structural transition, and fibre diffraction patterns are similar to those from the high-temperature form of Pf1. An α-helical polyalanine model of Pf3 based on the Pf1 structure was used to separate and phase the Bessel terms, which were then used to calculate $(5F_{\rm o} - 4F_{\rm c})$ maps. These maps were used to align and position the polypeptide chain, and the resulting model was refined by molecular dynamics (Welsh *et al.*, 1998).

The R-type bacterial flagellar filament structure (that has a very high molecular weight subunit) has been determined at 9 Å resolution by X-ray fibre diffraction (Yamashita *et al.*, 1998). Accurate intensities were taken from high-quality X-ray diffraction patterns and combined with phases obtained from electron cryomicroscopy, and solvent levelling was used to refine the phases.

Some studies of muscle provide a good example of the use of low-resolution fibre diffraction data, coupled with high-resolution crystal structures of some of the component molecules, to determine the structure of a complex. Holmes *et al.* (1990) constructed a model of F-actin based on the crystal structure of

the monomer, G-actin, and 8 Å fibre diffraction data, by either treating the monomer as a rigid body or dividing it into four separate rigid domains, and using a search procedure followed by least-squares refinement. The results gave the orientation of the actin monomer in the actin helix. This structure has since been refined using a genetic algorithm (Lorenz *et al.*, 1993) and normal-mode analysis (Tirion *et al.*, 1995). The genetic algorithm involved a Monte Carlo method of selecting subdomains to be refined and nonlinear least squares to obtain the best fit for the selected domains. In the normal-mode analysis, the model was parameterized in terms of its low-frequency vibrational modes to allow low-energy conformational changes and reduce the number of parameters which were optimized against the fibre diffraction data using nonlinear least squares.

Squire *et al.* (1993) have refined a low-resolution model of the muscle thin-filament structure that consists of four spheres representing each of the F-actin monomer subdomains and five spheres (fixed relative to each other) representing tropomyosin. Steric restraints were placed on the actin subdomain and thin-filament structures. The positions of the actin subdomains and the orientation of the tropomyosin were refined using a search procedure against fibre diffraction data from both 'resting' and 'activated' muscle at 25 Å resolution. More recent work has used a low-resolution model of the myosin head (based on the single-crystal atomic structure), a search procedure and simulated-annealing refinements to study myosin head configuration (Hudson *et al.*, 1997) and myosin rod packing (Squire *et al.*, 1998).

4.5.2.6.8. Reliability

As with structure determination in any area of crystallography, assessment of the reliability or precision of a structure is critically important. The most commonly used measure of reliability in fibre diffraction is the *R* factor, calculated as

$$R = \frac{\sum_i ||F|_i^{\rm o} - |F|_i^{\rm c}|}{\sum_i |F|_i^{\rm o}}, \qquad (4.5.2.74)$$

where $|F|_i^{\rm o}$ and $|F|_i^{\rm c}$ denote the observed (measured) and calculated, respectively, amplitude of either the samples (along *R*) of the cylindrically averaged intensity $I_l^{1/2}(R)$ (for a noncrystalline specimen) or the cylindrically averaged structure factors $I_l^{1/2}(R_{hk})$ (for a polycrystalline specimen). One way of assessing the significance of the *R* factor obtained in a particular structure determination is by comparing it with the 'largest likely *R* factor' (Wilson, 1950), *i.e.* the expected value of the *R* factor for a random distribution of atoms. Wilson (1950) showed that the largest likely *R* factor is 0.83 for a centric crystal and 0.59 for an acentric crystal. Although it does not provide a quantitative measure of structural reliability, the largest likely *R* factor does provide a useful yardstick for evaluating the significance of *R* factors obtained in structure determinations.

The largest likely *R* factor for fibre diffraction can be calculated from the amplitude statistics, which depend on the number of degrees of freedom, *m*, in the measured intensity (Stubbs, 1989; Millane, 1990*a*). Making use of these statistics shows that the largest likely *R* factor, R_m, for *m* components is given by (Stubbs, 1989; Millane, 1989*a*)

$$R_m = 2 - 2^{2-m} m \binom{2m-1}{m} B_{1/2}\left(\frac{m+1}{2}, \frac{m}{2}\right), \qquad (4.5.2.75)$$

where $\binom{m}{n}$ is the binomial coefficient and $B_x(m, n)$ the incomplete beta function. The beta function in equation (4.5.2.75) can be replaced by a finite series that is easy to evaluate (Millane, 1989*a*). The expression in equation (4.5.2.75) for R_m can be written in various approximate forms (Millane, 1990*d*, 1992*a*), the simplest being

$$R_m \simeq (2/\pi m)^{1/2} \qquad (4.5.2.76)$$

(Millane, 1990*d*), which shows that the largest likely *R* factor falls off approximately as $m^{-1/2}$ with increasing *m*. This is because it is easier to match the sum of a number of structure amplitudes than to match each of them individually. The important conclusion is that the largest likely *R* factor is smaller in fibre diffraction than in conventional crystallography (where $m = 1$ or 2), and it is smaller when there are more overlapping reflections. This means that for equivalent precision, the *R* factor must be smaller for a structure determined by fibre diffraction than for one determined by conventional crystallography. How much smaller depends on the number of overlapping reflections on the diffraction pattern.

In a structure determination, the data have different values of *m* at different positions on the diffraction pattern. Using the definition of the *R* factor, equation (4.5.2.74), shows that the largest likely *R* factor for a structure determination is given by (Millane, 1989*b*)

$$R = \frac{\sum_m N_m R_m S_m}{\sum_m N_m S_m}, \qquad (4.5.2.77)$$

where the sums are over the values of *m* on the diffraction pattern, N_m is the number of data that have *m* components, R_m is given by equation (4.5.2.75) and S_m is given by

$$S_m = \frac{\Gamma\big((m/2) + (1/2)\big)}{\Gamma(m/2)}, \qquad (4.5.2.78)$$

where $\Gamma(\cdot)$ is the gamma function. The quantities on the right-hand side of equation (4.5.2.77) are easily determined for a particular data set. The largest likely *R* factor decreases (since *m* increases) with increasing resolution of the data, increasing diameter of the molecule and decreasing order *u* of the helix symmetry. For example, for TMV at 5 Å resolution the largest likely *R* factor is 0.37, and at 3 Å resolution it is 0.31, whereas for a tenfold nucleic acid structure at 3 Å resolution it is 0.40 (Millane, 1989*b*, 1992*b*). This underlines the importance of comparing *R* factors obtained in a fibre diffraction analysis with the largest likely *R* factor; an *R* factor of 0.25 that may indicate a good protein structure may, or may not, indicate a well determined fibre structure.

Using approximations for R_m, S_m and *m* allows the following approximation for the largest likely *R* factor for a noncrystalline fibre to be derived (Millane, 1992*b*):

$$R \simeq 0.261(u d_{\rm max}/r_{\rm max})^{1/2}, \qquad (4.5.2.79)$$

where $d_{\rm max}$ is the resolution of the data. The approximation (4.5.2.79) is generally not good enough for calculating accurate largest likely *R* factors, but it does show the general behaviour with helix symmetry, molecular diameter and diffraction-data resolution. Other approximations to largest likely *R* factors have been derived that are quite accurate and also include the effect of a minimum resolution for the data (Millane, 1992*b*).

Largest likely *R* factors in fibre diffraction studies are typically between about 0.3 and 0.5, depending on the particular structure (Millane, 1989*b*, 1992*b*; Millane & Stubbs, 1992). Although the largest likely *R* factor does not give a quantitative assessment of the significance of an *R* factor obtained in a particular structure determination, it can be used as a guide to the significance. *R* factors obtained for well determined protein structures are

typically between about one-third and one-half of the corresponding largest likely R factor, depending on the resolution. It is therefore reasonable to expect the R factor for a well determined fibre structure to be between one-third and one-half of the largest likely R factor calculated for the structure. R factors should, therefore, generally be less than 0.15 to 0.25, depending on the particular structure and the resolution as illustrated by the examples presented in Millane & Stubbs (1992).

The free R factor (Brünger, 1997) has become popular in single-crystal crystallography as a tool for validation of refinements. The free R factor is more difficult to implement (but is probably even more important) in fibre diffraction studies because of the smaller data sets, but has been used to advantage in recent studies (Hudson *et al.*, 1997; Welsh *et al.*, 1998, 2000).

4.5.3. Electron crystallography of polymers

By D. L. Dorset

4.5.3.1. Is polymer electron crystallography possible?

As a crystallographic tool, the electron microscope has also made an important impact in polymer science. Historically, single-crystal electron diffraction information has been very useful for the interpretation of cylindrically averaged fibre X-ray patterns (Atkins, 1989), particularly when there is an extensive overlap of diffracted intensities. An electron diffraction pattern aids indexing of the fibre pattern and facilitates measurement of unit-cell constants, and the observation of undistorted plane-group symmetry similarly places important constraints on the identification of the space group (Geil, 1963; Wunderlich, 1973).

The concept of using electron diffraction intensities by themselves for the quantitative determination of crystal structures of polymers or other organics often has been met with scepticism (Lipson & Cochran, 1966). Difficulties experienced in the quantitative interpretation of images and diffraction intensities from 'hard' materials composed of heavy atoms (Hirsch *et al.*, 1965; Cowley, 1981), for example, has adversely affected the outlook for polymer structure analysis, irrespective of whether these reservations are important or not for 'soft' materials comprising light atoms. Despite the still commonly held opinion that no new crystal structures will be determined that are solely based on data collected in the electron microscope, it can be shown that this extremely pessimistic outlook is unwarranted. With proper control of crystallization (*i.e.* crystal thickness) and data collection, the electron microscope can be used quite productively for the direct determination of macromolecular structures at atomic resolution, not only to verify some of the previous findings of fibre X-ray diffraction analysis, but, more importantly, to determine new structures, even of crystalline forms that cannot be studied conveniently by X-rays as drawn fibres (Dorset, 1995*b*). The potential advantages of electron crystallography are therefore clear. The great advantage in scattering cross section of matter for electrons over X-rays permits much smaller samples to be examined by electron diffraction as single-crystalline preparations (Vainshtein, 1964). (Typical dimensions are given below.)

Electron crystallography can be defined as the quantitative use of electron micrographs and electron diffraction intensities for the determination of crystal structures. In the electron microscope, an electron beam illuminates a semitransparent object and the microscope objective lens produces an enlarged representation of the object as an image. If the specimen is thin enough and/or the electron energy is high enough, the weak-phase-object or 'kinematical' approximation is valid (Cowley, 1981), see Chapter 2.5. That is to say, there is an approximate one-to-one mapping of density points between the object mass distribution and the image, within the resolution limits of the instrument (as set by the objective lens aberrations and electron wavelength). The spatial relationships between diffraction and image planes of an electron microscope objective lens are reciprocal and related by Fourier transform operations (Cowley, 1988). While it is easy to transform from the image to the diffraction pattern, the reverse Fourier transform of the diffraction pattern to a high-resolution image requires solution of the famous crystallographic phase problem (as discussed for electron diffraction in Section 2.5.8).

Certainly, in electron diffraction studies, one must still be cognizant of the limitations imposed by the underlying scattering theory. An approximate 'quasi-kinematical' data set is often sufficient for the analysis (Dorset, 1995*a*). However (Dorset, 1995*b*), there are other important perturbations to diffraction intensities which should be minimized. For example, the effects of radiation damage while recording a high-resolution image are minimized by so-called 'low-dose' procedures (Tsuji, 1989).

4.5.3.2. Crystallization and data collection

The success of electron crystallographic determinations relies on the possibility of collecting data from *thin* single microcrystals. These can be grown by several methods, including self-seeding, epitaxic orientation, *in situ* polymerization on a substrate, in a Langmuir–Blodgett layer, *in situ* polymerization within a thin layer and polymerization in dilute solution. If these preparations do not provide sufficient information, then data can also be collected from microfibres. Thin cast films have also been examined after stretching.

Self-seeding (Blundell *et al.*, 1966) has been one of the most important techniques for growing single chain-folded lamellae. The technique is very simple. A dilute suspension of the polymer is made in a poor solvent. The temperature is raised to cause total solubilization of the macromolecule and then lowered to room temperature to crystallize ill-formed particles (mostly dendrites). The temperature is then elevated again until the suspension just clears, leaving small seeds of the polymer crystals behind. Upon lowering to a suitable temperature above ambience, which is then fixed, isothermal crystallization of well formed lamellae is allowed to occur over time. When the crystallization procedure is complete, the suspension can be cooled again to room temperature and the lamellae harvested. These lamellae are typically less than 10 nm thick, with lateral dimensions between 1.0 and 10.0 μm.

Epitaxic orientation techniques, to give alternative projections of the chain packing, have become increasingly important in recent years. While inorganic substrates have been described (Mauritz *et al.*, 1978), the use of organic layers for this purpose (Wittmann & Lotz, 1990; Lotz & Wittmann, 1993) has been more promising because these substrates are less easily contaminated by adsorbed gases and water vapour, and because the nucleation is anisotropic. Often the crystallization can be carried out from a cooled co-melt, *i.e.* a dilute solution of the polymer in the organic small molecule. When the liquidus curve of the eutectic phase diagram is crossed, the diluent crystals form first. Since these have a surface lattice spacing closely resembling that of the polymer-chain packing, the polymer chains can be directed to lie *along* the substrate surface, rather than normal to it, as the solidus line of the phase diagram is crossed. The substrate can then be removed by some suitable technique (sublimation, selective solvation) to permit the investigation of the oriented film. Variations of this procedure include crystallization of polymer-chain segments from the vapour phase onto a substrate (Wittmann & Lotz, 1985) and *in situ* crystallization of monomers that have first been epitaxically oriented on a suitable substrate (Rickert *et al.*, 1979).

A number of other possibilities for crystal growth also exist. Langmuir troughs have been used to orient monomers that may have hydrophilic moieties. If the monomers contain triple bonds that can be cross-linked, then a polymer film can be formed, *e.g.*,

if the condensed monomer film is exposed to ultraviolet light (Day & Lando, 1980). It may be possible to carry out the polymerization within a confined space (Rybnikar *et al.*, 1994) or in dilute solution (Liu & Geil, 1993) to form crystals suitable for electron diffraction data collection. In the latter case, whisker formation with the chain axis parallel to the lath plane has been observed. Films can be cast on a water surface by evaporation of an organic solvent from a polymer solution. The film can then be stretched to give a suitably oriented specimen for data collection (Vainshtein & Tatarinova, 1967). In addition, it may just be possible to obtain suitable data from drawn microfibres to supplement the single-crystal diffraction information from other preparations.

Data collection from these thin microcrystals often employs the selected-area diffraction technique in the electron microscope that is described in detail elsewhere (Dorset, 1995*b*). Using an approximately eucentric goniometric tilting device in the electron microscope, the sampling of three-dimensional reciprocal space is tomographic, *i.e.* the tilts of a nearly planar Ewald sphere surface (owing to the very small electron wavelength) are always referred to a set of reciprocal axes that intersect (0, 0, 0). For any given crystal habit, there is always a missing set of data owing to the physical limitation to the tilt imposed by the finite thickness of the specimen holder within the pole-piece gap of the electron microscope objective lens (Vainshtein, 1964). For this reason, it is desirable to crystallize two orthogonal orientations of the chain packing (using the above-mentioned approaches), if possible, so that all of the reciprocal lattice can be sampled. If electron micrographs are to be used as an additional source of crystallographic phases then 'low-dose' techniques for recording such images should be employed to reduce the deleterious effects of radiation damage caused by the inelastic interactions of the electron beam with the crystalline sample (Tsuji, 1989).

When the diffraction patterns are recorded on photographic film and these are then measured with a densitometer, relative reflection intensities can often be expressed simply as the integrated peak area without need for a Lorentz correction (Dorset, 1995*b*). Only if the diffraction maxima are extensively arced (*e.g.* in patterns from epitaxic films) is such a correction required. That is to say, $|\Phi_{\rm obs}| \propto KI_{\rm obs}^{1/2}$ where $|\Phi_{\rm obs}|$ is the observed structure-factor magnitude. Assuming the kinematical approximation holds, the calculated value is

$$\Phi_h^{\rm calc} = \sum_i f_i \exp 2\pi i(\mathbf{h}\cdot\mathbf{r}),$$

where f_i are the electron scattering factors (Doyle & Turner, 1968), *e.g.* as tabulated in Table 4.3.1.1 in *IT* C. By analogy with X-ray crystallography (see Chapter 2.2), normalized values can be found from

$$|E_h|^2 = I_h^{\rm obs}/\varepsilon\sum_i f_i^2,$$

with the usual scaling condition that $\langle E_h^2\rangle = 1.000$. [Note, however, that these intensities only describe the chain monomer packing in the 'stem' region of the lamellar microcrystal. Details owing to the surface chain folds are lost (even if they are strictly periodic), because of reasons similar to those described by Cowley (1961) for the electron scattering from elastically bent silicate crystals.]

4.5.3.3. Crystal structure analysis

Two approaches to crystal structure analysis are generally employed in polymer electron crystallography. As already mentioned, the procedure adapted from fibre X-ray crystallography relies on the construction of a model (Brisse, 1989; Perez & Chanzy, 1989). Conformational searches (Campbell Smith & Arnott, 1978) simultaneously minimize the fit of observed diffraction data to calculated values (the *R* factor based on structure factors computed *via* known atomic scattering factors) and a nonbonded atom–atom potential function (Tadokoro, 1979). Reviews of structures solved by this approach have been published (Dorset, 1989, 1995*b*).

Recently, direct phasing methods of the kind used in X-ray crystallography (Chapter 2.2 and, applied to electron diffraction, Section 2.5.8) have also been found to be particularly effective for electron crystallographic structure analyses (Dorset, 1995*b*). While the Fourier transform of an electron micrograph would be the most easily imagined direct method, yielding crystallographic phases after image analysis (see Section 2.5.5), this use of micrographs has been of less importance to polymer crystallography than it has been in the study of globular proteins, even though there is at least one notable example where it has been helpful (Isoda *et al.*, 1983*a*) for the determination of a structure from X-ray fibre data. On the other hand, high-resolution images of polymer crystals are of considerable use for the characterization of packing defects (Isoda *et al.*, 1983*b*).

In polymer electron crystallography, the sole reliance on the diffraction intensities for structure analysis has proven, in recent years, to be quite effective. Several direct-methods approaches have been pursued, including the use of probabilistic techniques, either in the symbolic addition procedure, or in more automated procedures involving the tangent formula (see Chapter 2.2). The Sayre (1952) equation has been found to be particularly effective, where the correct structure is identified *via* some figure of merit after algebraic phase values are used to generate multiple solutions (Stanley, 1986). More recently, maximum-entropy and likelihood methods (Gilmore *et al.*, 1993) have also been effective for solving such structures. After the initial atomic model is found, it can be improved by refinement, generally using Fourier techniques. Least-squares refinement can be carried out under most favourable circumstances (Dorset, 1995*a*), but requires the availability of a sufficient number of diffraction data. Even so, the refinement of thermal parameters must be uncoupled from that of the atomic positions. Also, positional shifts must be dampened (if X-ray crystallographic software is used) to prevent finding a false minimum, especially if the kinematical *R* factor is used as a figure of merit.

4.5.3.4. Examples of crystal structure analyses

At least four kinds of electron diffraction intensity data sets have been used as tests for direct phase determination *via* the approaches mentioned above.

Case 1: Zonal data sets – view down the chain axis. Such data are from the least optimal projection of the polymer packing, because of extensive atomic overlap along the chain axis. Initially, symbolic addition was used to find phase values for *hk*0 data sets from six representative polymers, including three complicated saccharide structures (Dorset, 1992). Most of the determinations were strikingly successful. Later, an unknown data set from the polysaccharide chitosan was obtained from Grenoble (Mazeau *et al.*, 1994) and direct phase determination was able to find a correct model (Dorset, 1995*b*). More recently, other polymers have been tested [including one case where an electron micrograph provided many of the starting phase terms (Dorset, 1995*b*)] also comparing favourably with the solution found by energy minimization of a linkage model. For all examples considered so far, the projected symmetry was centrosymmetric.

Later, it was found that a partial phase set provided by symbolic addition could be expanded to the complete zone by the Sayre equation (Dorset *et al.*, 1995). In all of these tests (Dorset, 1995*b*), there were only one or two examples where there were serious deviations from the phase terms found by other methods. Even in these instances, the potential maps could still be used as

Table 4.5.3.1. *Structure analysis of poly-γ-methyl-L-glutamate in the β form*

h0l	$\|E_h\|$	$\|F_o\|$	$\|F_c\|$	φ (°) (previous)	φ (°) (this study)
002	0.48	0.72	0.57	−63	−51
004	0.43	0.38	0.31	49	73
006	3.01	1.47	0.88	1	−3
100	1.48	2.12	2.37	0	0
200	1.03	1.04	1.06	0	0
300	0.30	0.65	0.89	0	0
400	0.35	0.15	0.46	0	0
500	0.23	0.07	0.04	180	180
101	0.75	1.02	0.67	−169	−178
201	0.32	0.31	0.42	90	108
102	0.42	0.48	0.56	17	14
202	0.40	0.33	0.64	41	43
103	0.95	0.85	0.77	88	90
203	0.51	0.36	0.42	91	88
303	0.12	0.06	0.31	92	87
403	0.13	0.04	0.54	90	90
104	0.66	0.45	0.27	−22	−13
105	0.55	0.28	0.29	−26	−7
106	1.75	0.69	0.58	5	−5

Fractional coordinates

	This study		Vainshtein & Tatarinova (1967)	
	x	*z*	*x*	*z*
Cα, β	0.048	0.000	0.042	0.000
C′	0.067	0.331	0.092	0.330
O	0.281	0.335	0.300	0.330
N	0.000	0.161	−0.025	0.175

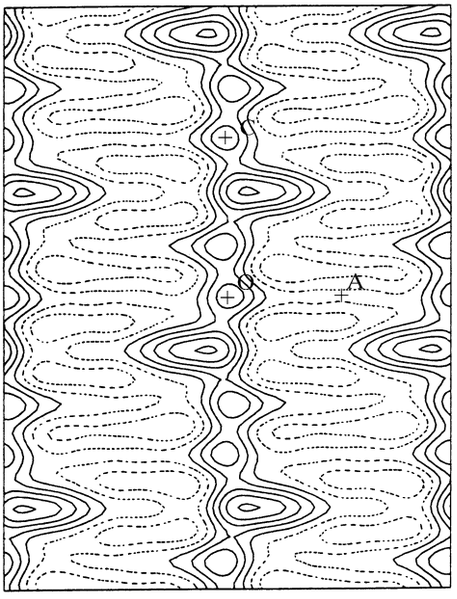


Fig. 4.5.3.1. Initial potential map for poly-γ-methyl-L-glutamate (plane group *pg*) found with phases generated by the Sayre equation.

envelopes for the actual projection of the chain structure (Dorset, 1992).

Case 2: Zonal data set – view onto the chain axes. Electron diffraction data from a projection onto the polymer chain axes would be more useful if individual atomic positions were to be resolved. An interesting example where such a view can be obtained is an *h0l* data set from the polypeptide poly-γ-methyl-L-glutamate. Electron diffraction data were collected from stretched films by Vainshtein & Tatarinova (1967). In projection, the cell constants are $a = 4.72$, $c = 6.83$ Å with plane-group symmetry *pg*. As shown in Table 4.5.3.1, there were 19 unique intensity data used for the analysis. After initial phase assignment by symbolic addition, a correct solution could be visualized which, after Fourier refinement (Dorset, 1995*b*), differed from the original one by a mean phase difference of only 6°.

The progress of this structure analysis can be reviewed to give a representative example. Since the *h*00 reflections have centrosymmetric phases, the value $\varphi_{100} = 0$ was chosen as a single origin-defining point. From high-probability Σ_1 three-phase invariants (assessed after calculation of normalized structure factors $|E_h|$), one could assign $\varphi_{200} = \varphi_{400} = 0$. Symbolic values were then given to three other phases, *viz.* $\varphi_{106} = a$; $\varphi_{103} = b$; $\varphi_{101} = c$. From this entire basis, other values could be found from highly probable Σ_2 three-phase invariants, as follows:

$$\begin{aligned}
\varphi_{006} &= \varphi_{106} + \varphi_{\bar{1}00} \;\;\therefore\;\; \varphi_{006} = a \\
\varphi_{105} &= \varphi_{006} + \varphi_{10\bar{1}} \;\;\therefore\;\; \varphi_{105} = a - c + \pi \\
\varphi_{203} &= \varphi_{106} + \varphi_{10\bar{3}} \;\;\therefore\;\; \varphi_{203} = a - b + \pi \\
\varphi_{300} &= \varphi_{100} + \varphi_{200} \;\;\therefore\;\; \varphi_{300} = 0 \\
\varphi_{002} &= \varphi_{103} + \varphi_{\bar{1}01} \;\;\therefore\;\; \varphi_{002} = b - c \\
\varphi_{004} &= \varphi_{006} + \varphi_{00\bar{2}} \;\;\therefore\;\; \varphi_{004} = a - b + c.
\end{aligned}$$

(These invariant relationships include phase interactions among symmetry-related Miller indices characteristic of the plane group.) Additionally $c = \pi$ could be specified to complete origin definition for the zone. It was then possible to permute values of *a* and *b* to arrive at test phase values for this subset, *i.e.* to generate a multiple set of solutions. When $a = 0$, $b = \pi/2$, the map in Fig. 4.5.3.1 was observed. After finding trial atomic positions for Fourier refinement (assuming that two carbon-atom positions were eclipsed in this projection), the final phase set was found as shown in Table 4.5.3.1. Although the crystallographic residual to the observed data, calculated with the model coordinates, was rather large (0.32), there was a close agreement with the earlier determination.

More recently a similar data set, collected from oriented crystal 'whiskers' of poly(*p*-oxybenzoate) in plane group *pg* was analysed. Again the Sayre equation, *via* a multisolution approach, was used to produce a map that contained 13 of 18 possible atomic positions for the two subunits in the asymmetric unit. The complete structure was observed after the remaining five atom sites were identified in two subsequent cycles of Fourier refinement (Liu *et al.*, 1997) and the average atomic positions were found to be within 0.2 Å of the model derived from an energy minimization.

Case 3: Three-dimensional data – single crystal orientation. The first data set from a chain-folded lamella for a direct structure analysis was a centrosymmetric set (space group $P2_1/n$) from poly(1,4-*trans*-cyclohexanediyl dimethylene succinate), composed of 87 reflections (Brisse *et al.*, 1984). The phase determination was quite successful and atomic positions could be found as somewhat blurred density maxima in the three-dimensional maps (Dorset, 1991*a*). A model was constructed from these positions and the bonding parameters optimized to give the best fit to the data ($R = 0.29$).

Noncentrosymmetric three-dimensional intensity sets (orthorhombic space group $P2_12_12_1$) from the polysaccharides mannan (form I) (Chanzy *et al.*, 1987) and chitosan (Mazeau *et al.*, 1994) were also collected from tilted crystals. In both cases, direct phase determination by symbolic addition *via* an algebraic unknown was successful, even though the data were not sampled along the chain repeat. For the former polymer, a monomer model could be fitted to the blurred density profile, much as one would fit a polypeptide sequence to a continuous electron-density map (Dorset & McCourt, 1993; Dorset, 1995*c*). If the Sayre equation were used to predict phases and amplitudes within the 'missing cone' of unsampled data, then the fit of the monomer could be much more highly constrained.

Case 4: Three-dimensional data – two crystal orientations. The optimal case for collection of diffraction data is when two

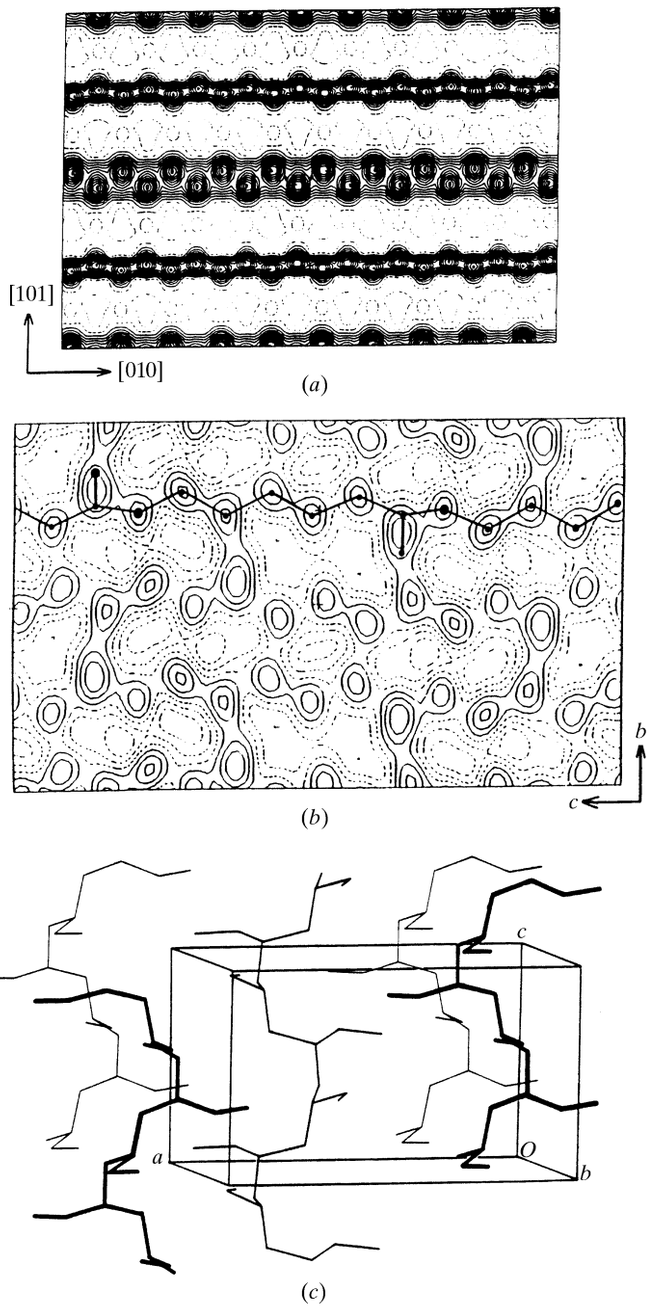


Fig. 4.5.3.2. Crystal structures of linear polymers determined from three-dimensional data. (*a*) Polyethylene; (*b*) poly(ε-caprolactone); (*c*) poly(1-butene), form (III).

orthogonal projections of the same polymer polymorph can be obtained, respectively, by self-seeding and epitaxic orientation. While tilting these specimens, all of reciprocal space can be sampled for intensity data collection.

Polyethylene crystals were used to collect 50 unique maxima (Hu & Dorset, 1989) and, *via* symbolic addition, the centrosymmetric phases of 40 reflections (space group *Pnma*) could be readily determined (Dorset, 1991*b*). The structural features were readily observed in the three-dimensional potential maps (Fig. 4.5.3.2*a*), and atomic coordinates (with estimated values for hydrogen-atom positions) could be refined by least squares (Dorset, 1995*b*) to give a final R value of 0.19.

Poly(ε-caprolactone) was epitaxically crystallized on benzoic acid and, with *hk*0 data from solution-crystallized samples, a unique set of 47 intensities was collected for the noncentrosymmetric orthorhombic unit cell (space group $P2_12_12_1$) (Hu & Dorset, 1990). Direct phase determination was achieved *via* symbolic addition, using one algebraic unknown to assign values to 30 reflections (Dorset, 1991*c*). Atomic positions along the chain repeat, including the carbonyl position, were clearly discerned in the [100] projection (Fig. 4.5.3.2*b*) and the three-dimensional model was constructed to fit to the map calculated from all phased data, yielding a final crystallographic residual R = 0.21. This independent determination was able to distinguish between two rival fibre X-ray structures, in favour of the one that predicted a nonplanar chain conformation. Because of the methylene repeat, this is actually a difficult structure to solve by automated techniques. For example, the tangent formula and *SnB* (Miller *et al.*, 1993) could only find chain zigzag positions and not the position of the carbonyl oxygen atom (Dorset, 1995*b*).

The most complicated complete polymer crystal structure solved so far by direct methods using electron diffraction data (Dorset *et al.*, 1994) was based on 125 unique data (space group $P2_12_12_1$) from isotactic poly(1-butene), form (III), using orthogonal molecular orientations crystallized in Strasbourg (Kopp *et al.*, 1994). Initially, the standard NQEST figure of merit (FOM) (De Titta *et al.*, 1975) was not suitable for identifying the correct solution among the multiple sets generated with the tangent formula. A solution could only be found when a separate phase determination was carried out with the *hk*0 data to compare with the multiple solutions generated. More recently, the minimal principle (Hauptman, 1993), used as a FOM with the tangent formula or with a multiple random structure generator, *SnB*, correctly identified the structure on the first try (Dorset, 1995*b*). The maps clearly show individual carbon-atom positions in a 4_1 helix that parallels 2_1 helices of the space group (Fig. 4.5.3.2*c*). After Fourier refinement, the crystallographic residual was R = 0.26. The previous powder X-ray diffraction determination was based on only 21 diffraction maxima, some of which had as many as 15 individual contributors.

The work of both RPM and DLD was supported by the US National Science Foundation (grant DBI-9722862 for RPM).

References

Alexeev, D. G., Lipanov, A. A. & Skuratovskii, I. Y. (1992). *Patterson methods in fibre diffraction. Int. J. Biol. Macromol.* **14**, 139–144.

Arnott, S. (1980). *Twenty years hard labor as a fibre diffractionist*. In *Fibre Diffraction Methods*, ACS Symposium Series, Vol. 141, edited by A. D. French & K. H. Gardner, pp. 1–30. Washington: American Chemical Society.

Arnott, S., Chandrasekaran, R., Millane, R. P. & Park, H. (1986). *DNA–RNA hybrid secondary structures. J. Mol. Biol.* **188**, 631–640.

Arnott, S. & Mitra, A. K. (1984). *X-ray diffraction analyses of glycosamionoglycans*. In *Molecular Biophysics of the Extracellular Matrix*, edited by S. Arnott, D. A. Rees & E. R. Morris, pp. 41–67. Clifton: Humana Press.

Arnott, S., Wilkins, M. H. F., Fuller, W. & Langridge, R. (1967). *Molecular and crystal structures of double-helical RNA III. An 11-fold molecular model and comparison of the agreement between the observed and calculated three-dimensional diffraction data for 10- and 11-fold models. J. Mol. Biol.* **27**, 535–548.

Arnott, S. & Wonacott, A. J. (1966). *The refinement of the crystal and molecular structures of polymers using X-ray data and stereochemical constraints. Polymer*, **7**, 157–166.

Atkins, E. D. T. (1989). *Crystal structure by X-ray diffraction*. In *Comprehensive Polymer Science*, Vol. 1. *Polymer Characterization*, edited by G. A. Allen, pp. 613–650. Oxford: Pergamon Press.

Barham, P. J. (1993). *Crystallization and morphology of semicrystalline polymers*. In *Materials Science and Technology. A Comprehensive Treatment*, Vol. 12. *Structure and Properties of Polymers*, edited by E. L. Thomas, pp. 153–212. Weinheim: VCH.

Baskaran, S. & Millane, R. P. (1999*a*). *Bayesian image reconstruction from partial image and aliased spectral intensity data. IEEE Trans. Image Process.* **8**, 1420–1434.

Baskaran, S. & Millane, R. P. (1999*b*). *Model bias in Bayesian image reconstruction from X-ray fiber diffraction data. J. Opt. Soc. Am. A*, **16**, 236–245.

Biswas, A. & Blackwell, J. (1988*a*). *Three-dimensional structure of main-chain liquid-crystalline copolymers. 1. Cylindrically averaged intensity transforms of single chains. Macromolecules*, **21**, 3146–3151.

Biswas, A. & Blackwell, J. (1988*b*). *Three-dimensional structure of main-chain liquid-crystalline copolymers. 2. Interchain interference effects. Macromolecules*, **21**, 3152–3158.

Biswas, A. & Blackwell, J. (1988*c*). *Three-dimensional structure of main-chain liquid-crystalline copolymers. 3. Chain packing in the solid state. Macromolecules*, **21**, 3158–3164.

Blackwell, J., Gutierrez, G. A. & Chivers, R. A. (1984). *Diffraction by aperiodic polymer chains: the structure of liquid crystalline copolyesters. Macromolecules*, **17**, 1219–1224.

Blundell, D. J., Keller, A. & Kovacs, A. J. (1966). *A new self-nucleation phenomenon and its application to the growing of polymer crystals from solution. J. Polym. Sci. Polym. Lett. Ed.* **4**, 481–486.

Brisse, F. (1989). *Electron diffraction of synthetic polymers: the model compound approach to polymer structure. J. Electron Microsc. Tech.* **11**, 272–279.

Brisse, F., Remillard, B. & Chanzy, H. (1984). *Poly(1,4-trans-cyclohexanediyldimethylene succinate). A structural determination using X-ray and electron diffraction. Macromolecules*, **17**, 1980–1987.

Brünger, A. T. (1992). *X-PLOR*. Version 3.1. New Haven: Yale University Press.

Brünger, A. T. (1997). *Free R value: cross-validation in crystallography. Methods Enzymol.* **277**, 366–396.

Cael, J. J., Winter, W. T. & Arnott, S. (1978). *Calcium chondroitin 4-sulfate: molecular conformation and organization of polysaccharide chains in a proteoglycan. J. Mol. Biol.* **125**, 21–42.

Campbell Smith, P. J. & Arnott, S. (1978). *LALS: a linked-atom least-squares reciprocal-space refinement system incorporating stereochemical constraints to supplement sparse diffraction data. Acta Cryst.* A**34**, 3–11.

Chandrasekaran, R. & Arnott, S. (1989). *The structures of DNA and RNA helices in oriented fibres*. In *Landolt–Bornstein Numerical Data and Functional Relationships in Science and Technology*, Vol. VII/1b, edited by W. Saenger, pp. 31–170. Berlin, Heidelberg: Springer-Verlag.

Chandrasekaran, R., Radha, A. & Lee, E. J. (1994). *Structural roles of calcium ions and side chains in welan: an X-ray study. Carbohydr. Res.* **252**, 183–207.

Chanzy, H., Perez, S., Miller, D. P., Paradossi, G. & Winter, W. T. (1987). *An electron diffraction study of mannan I. Crystal and molecular structure. Macromolecules,* **20**, 2407–2413.

Chivers, R. A. & Blackwell, J. (1985). *Three-dimensional structure of copolymers of p-hydroxybenzoic acid and 2-hydroxy-6-naphthoic acid: a model for diffraction from a nematic structure. Polymer*, **26**, 997–1002.

Clark, E. S. & Muus, I. T. (1962). *The relationship between Bragg reflections and disorder in crystalline polymers. Z. Kristallogr.* **117**, 108–118.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *The structure of synthetic polypeptides. I. The transform of atoms on a helix. Acta Cryst.* **5**, 581–586.

Cowley, J. M. (1961). *Diffraction intensities from bent crystals. Acta Cryst.* **14**, 920–927.

Cowley, J. M. (1981). *Diffraction Physics*, 2nd revised ed. Amsterdam: North-Holland.

Cowley, J. M. (1988). *Imaging [and] imaging theory*. In *High-Resolution Transmission Electron Microscopy and Associated Techniques*, edited by P. Busek, J. M. Cowley & L. Eyring, pp. 3–57. New York: Oxford University Press.

Crowther, R. A., DeRosier, D. J. & Klug, A. (1970). *The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc. R. Soc. London Ser. A*, **317**, 319–344.

Daubeny, R. de P., Bunn, C. W. & Brown, C. J. (1954). *The crystal structure of polyethylene terephthalate. Proc. R. Soc. London Ser. A*, **226**, 531–542.

Day, D. & Lando, J. B. (1980). *Structure determination of a poly(diacetylene) monolayer. Macromolecules*, **13**, 1483–1487.

De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Use of negative quartet cosine invariants as a phasing figure of merit: NQEST. Acta Cryst.* A**31**, 472–479.

Dorset, D. L. (1989). *Electron diffraction from crystalline polymers*. In *Comprehensive Polymer Science*, Vol. 1. *Polymer Characterization*, edited by G. A. Allen, pp. 651–668. Oxford: Pergamon Press.

Dorset, D. L. (1991*a*). *Is electron crystallography possible? The direct determination of organic crystal structures. Ultramicroscopy*, **38**, 23–40.

Dorset, D. L. (1991*b*). *Electron diffraction structure analysis of polyethylene. A direct phase determination. Macromolecules*, **24**, 1175–1178.

Dorset, D. L. (1991*c*). *Electron crystallography of linear polymers: direct structure analysis of poly(ε-caprolactone). Proc. Natl Acad. Sci. USA*, **88**, 5499–5502.

Dorset, D. L. (1992). *Electron crystallography of linear polymers: direct phase determination for zonal data sets. Macromolecules*, **25**, 4425–4430.

Dorset, D. L. (1995*a*). *Comments on the validity of the direct phasing and Fourier methods in electron crystallography. Acta Cryst.* A**51**, 869–879.

Dorset, D. L. (1995*b*). *Structural Electron Crystallography*. New York: Plenum.

Dorset, D. L. (1995*c*). *Filling the cone – overcoming the goniometric tilt limit in electron crystallography by direct methods. Am. Cryst. Assoc. Abstr.* Series 2, 23, p. 89.

Dorset, D. L., Kopp, S., Fryer, J. R. & Tivol, W. T. (1995). *The Sayre equation in electron crystallography. Ultramicroscopy*, **57**, 59–89.

Dorset, D. L. & McCourt, M. P. (1993). *Electron crystallographic analysis of a polysaccharide structure – direct phase determination and model refinement for mannan I. J. Struct. Biol.* **111**, 118–124.

Dorset, D. L., McCourt, M. P., Kopp, S., Wittmann, J.-C. & Lotz, B. (1994). *Direct determination of polymer crystal structures by electron crystallography – isotactic poly(1-butene), form (III). Acta Cryst.* B**50**, 201–208.

Doyle, P. A. & Turner, P. S. (1968). *Relativistic Hartree–Fock X-ray and electron scattering factors. Acta Cryst.* A**24**, 390–397.

Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. New York: Springer-Verlag.

Finkenstadt, V. L. & Millane, R. P. (1998). *Fiber diffraction patterns for general unit cells: the cylindrically projected reciprocal lattice. Acta Cryst.* A**54**, 240–248.

Forsyth, V. T., Mahendrasingam, A., Pigram, W. J., Greenall, R. J., Bellamy, K., Fuller, W. & Mason, S. A. (1989). *Neutron fibre diffraction study of DNA hydration. Int. J. Biol. Macromol.* **11**, 236–240.

Franklin, R. E. (1955). *Structure of tobacco mosaic virus. Nature (London)*, **175**, 379–381.

Franklin, R. E. & Gosling, R. G. (1953). *The structure of sodium thymonucleate fibres. II. The cylindrically symmetrical Patterson function. Acta Cryst.* **6**, 678–685.

Franklin, R. E. & Holmes, K. C. (1958). *Tobacco mosaic virus: application of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution. Acta Cryst.* **11**, 213–220.

Franklin, R. E. & Klug, A. (1955). *The splitting of layer lines in X-ray fibre diagrams of helical structures: application to tobacco mosaic virus. Acta Cryst.* **8**, 777–780.

Fraser, R. D. B. & MacRae, T. P. (1973). *Conformations in Fibrous Proteins*. New York: Academic Press.

Fraser, R. D. B., MacRae, T. P., Miller, A. & Rowlands, R. J. (1976). *Digital processing of fibre diffraction patterns. J. Appl. Cryst.* **9**, 81–94.

Fraser, R. D. B., Suzuki, E. & MacRae, T. P. (1984). *Computer analysis of X-ray diffraction patterns*. In *Structure of Crystalline Polymers*, edited by I. H. Hall, pp. 1–37. New York: Elsevier.

French, A. D. & Gardner, K. H. (1980). Editors. *Fibre Diffraction Methods*. ACS Symposium Series, Vol. 141. Washington: American Chemical Society.

Geil, P. H. (1963). *Polymer Single Crystals.* New York: John Wiley & Sons.

Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Application of the maximum entropy method to powder diffraction and electron crystallography. Proc. R. Soc. London Ser. A*, **442**, 97–111.

Gonzalez, A., Nave, C. & Marvin, D. A. (1995). *Pf1 filamentous bacteriophage: refinement of a molecular model by simulated annealing using 3.3 Å resolution X-ray fiber diffraction data. Acta Cryst.* D**51**, 792–804.

Graaf, H. de (1989). *On the calculation of small-angle diffraction patterns from distorted lattices. Acta Cryst.* A**45**, 861–870.

Grubb, D. T. (1993). *Elastic properties of crystalline polymers*. In *Materials Science and Technology. A Comprehensive Treatment*, Vol. 12. *Structure and Properties of Polymers*, edited by E. L. Thomas, pp. 301–356. Weinheim: VCH.

Hall, I. H. (1984). Editor. *Structure of Crystalline Polymers*. New York: Elsevier.

Hall, I. H., Neisser, J. Z. & Elder, M. (1987). *A computer-based method of measuring the integrated intensities of the reflections on the X-ray diffraction photograph of an oriented crystalline polymer. J. Appl. Cryst.* **20**, 246–255.

Hamilton, W. C. (1965). *Significance tests on the crystallographic R factor. Acta Cryst.* **18**, 502–510.

Hauptman, H. A. (1993). *A minimal principle in X-ray crystallography: starting in a small way. Proc. R. Soc. London Ser. A*, **442**, 3–12.

Hendricks, S. & Teller, E. (1942). *X-ray interference in partially ordered layer lattices. J. Chem. Phys.* **10**, 147–167.

Hirsch, P. B., Howie, A., Nicholson, P. B., Pashley, D. W. & Whelan, M. J. (1965). *Electron Microscopy of Thin Crystals.* London: Butterworths.

Hofmann, D., Schneider, A. I. & Blackwell, J. (1994). *Molecular modelling of the structure of a wholly aromatic thermotropic copolyester. Polymer*, **35**, 5603–5610.

Holmes, K. C. & Barrington Leigh, J. (1974). *The effect of disorientation on the intensity distribution of non-crystalline fibres. I. Theory. Acta Cryst.* A**30**, 635–638.

Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. (1990). *Atomic model of the actin filament. Nature (London)*, **347**, 44–49.

Holmes, K. C., Stubbs, G. J., Mandelkow, E. & Gallwitz, U. (1975). *Structure of tobacco mosaic virus at 6.7 Å resolution. Nature (London)*, **254**, 192–196.

Hosemann, R. & Bagchi, S. N. (1962). *Direct Analysis of Diffraction by Matter*. Amsterdam: North-Holland.

Hu, H. & Dorset, D. L. (1989). *Three-dimensional electron diffraction structure analysis of polyethylene. Acta Cryst.* B**45**, 283–290.

Hu, H. & Dorset, D. L. (1990). *Crystal structure of poly(ε-caprolactone). Macromolecules*, **23**, 4604–4607.

Hudson, L., Harford, J. J., Denny, R. C. & Squire, J. M. (1997). *Myosin head configuration in relaxed fish muscle: resting state myosin heads must swing axially by up to 150 Å or turn upside down to reach rigor. J. Mol. Biol.* **273**, 440–455.

Iannelli, P. (1994). *FWR: a computer program for refining the molecular structure in the crystalline phase of polymers based on the analysis of the whole X-ray fibre diffraction patterns. J. Appl. Cryst.* **27**, 1055–1060.

Isoda, S., Tsuji, M., Ohara, M., Kawaguchi, A. & Katayama, K. (1983*a*). *Structural analysis of β-form poly(p-xylene) starting from a high-resolution image. Polymer*, **24**, 1155–1161.

Isoda, S., Tsuji, M., Ohara, M., Kawaguchi, A. & Katayama, K. (1983*b*). *Direct observation of dislocations in polymer single crystals. Makromol. Chem. Rapid Commun.* **4**, 141–144.

Ivanova, M. I. & Makowski, L. (1998). *Iterative low-pass filtering for estimation of the background in fiber diffraction patterns. Acta Cryst.* A**54**, 626–631.

Klug, A., Crick, F. H. C. & Wyckoff, H. W. (1958). *Diffraction from helical structures. Acta Cryst.* **11**, 199–213.

Kopp, S., Wittmann, J. C. & Lotz, B. (1994). *Epitaxial crystallization and crystalline polymorphism of poly(1-butene): form (III) and (II). Polymer*, **35**, 908–915.

Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures*, p. 381. Ithaca: Cornell University Press.

Liu, J. & Geil, P. H. (1993). *Morphological observations of nascent poly(p-oxabenzoate). Polymer*, **34**, 1366–1374.

Liu, J., Yuan, B.-L., Geil, P. H. & Dorset, D. L. (1997). *Chain conformation and molecular packing in poly(p-oxybenzoate) single crystals at ambient temperature. Polymer*, **38**, 6031–6047.

Lobert, S., Heil, P. D., Namba, K. & Stubbs, G. (1987). *Preliminary X-ray fibre diffraction studies of cucumber green mottle mosaic virus, watermelon strain. J. Mol. Biol.* **196**, 935–938.

Lobert, S. & Stubbs, G. (1990). *Fibre diffraction analysis of cucumber green mottle mosaic virus using limited numbers of heavy-atom derivatives. Acta Cryst.* A**46**, 993–997.

Lorenz, M. & Holmes, K. C. (1993). *Computer processing and analysis of X-ray fibre diffraction data. J. Appl. Cryst.* **26**, 82–91.

Lorenz, M., Popp, D. & Holmes, K. C. (1993). *Refinement of the F-actin model against X-ray fibre diffraction data by the use of a directed mutation algorithm. J. Mol. Biol.* **234**, 826–836.

Lotz, B. & Wittmann, J. C. (1993). *Structure of polymer single crystals*. In *Materials Science and Technology. A Comprehensive Treatment*, Vol. 12. *Structure and Properties of Polymers*, edited by E. L. Thomas, pp. 79–154. Weinheim: VCH.

MacGillavry, C. H. & Bruins, E. M. (1948). *On the Patterson transforms of fibre diagrams. Acta Cryst.* **1**, 156–158.

Makowski, L. (1978). *Processing of X-ray diffraction data from partially oriented specimens. J. Appl. Cryst.* **11**, 273–283.

Makowski, L. (1982). *The use of continuous diffraction data as a phase constraint. II. Application to fibre diffraction data. J. Appl. Cryst.* **15**, 546–557.

Makowski, L., Caspar, D. L. D. & Marvin, D. A. (1980). *Filamentous bacteriophage Pf1 structure determined at 7 Å resolution by refinement of models for the α-helical subunit. J. Mol. Biol.* **140**, 149–181.

Mandelkern, L. (1989). *Crystallization and melting.* In *Comprehensive Polymer Science*, Vol. 2. *Polymer Properties*, edited by C. Booth & C. Price, pp. 363–414. Oxford: Pergamon Press.

Mandelkow, E. & Holmes, K. C. (1974). *The positions of the N-terminus and residue 68 in tobacco mosaic virus. J. Mol. Biol.* **87**, 265–273.

Mandelkow, E., Stubbs, G. & Warren, S. (1981). *Structures of the helical aggregates of tobacco mosaic virus protein. J. Mol. Biol.* **152**, 375–386.

Marvin, D. A., Bryan, R. K. & Nave, C. (1987). *Pf1 inovirus. Electron density distribution calculated by a maximum entropy algorithm from native fiber diffraction data to 3 Å resolution and single isomorphous replacement data to 5 Å resolution. J. Mol. Biol.* **193**, 315–343.

Mauritz, K. A., Baer, E. & Hopfinger, A. J. (1978). *The epitaxial crystallization of macromolecules. J. Polym. Sci. Macromol. Rev.* **13**, 1–61.

Mazeau, K., Winter, W. T. & Chanzy, H. (1994). *Molecular and crystal structure of a high-temperature polymorph of chitosan from electron diffraction data. Macromolecules*, **27**, 7606–7612.

Millane, R. P. (1988). *X-ray fibre diffraction.* In *Crystallographic Computing 4. Techniques and New Technologies*, edited by N. W. Isaacs & M. R. Taylor, pp. 169–186. Oxford University Press.

Millane, R. P. (1989*a*). *R factors in X-ray fibre diffraction. I. Largest likely R factors for N overlapping terms. Acta Cryst.* A**45**, 258–260.

Millane, R. P. (1989*b*). *R factors in X-ray fibre diffraction. II. Largest likely R factors. Acta Cryst.* A**45**, 573–576.

Millane, R. P. (1989*c*). *Relating reflection boundaries in X-ray fibre diffraction patterns to specimen morphology and their use for intensity measurement. J. Macromol. Sci. Phys.* B**28**, 149–166.

Millane, R. P. (1990*a*). *Intensity distributions in fibre diffraction. Acta Cryst.* A**46**, 552–559.

Millane, R. P. (1990*b*). *Phase retrieval in crystallography and optics. J. Opt. Soc. Am. A*, **7**, 394–411.

Millane, R. P. (1990*c*). *Polysaccharide structures: X-ray fibre diffraction studies*. In *Computer Modeling of Carbohydrate Molecules.* ACS Symposium Series No. 430, edited by A. D. French & J. W. Brady, pp. 315–331. Washington: American Chemical Society.

Millane, R. P. (1990*d*). *R factors in X-ray fibre diffraction. III. Asymptotic approximations to largest likely R factors. Acta Cryst.* A**46**, 68–72.

Millane, R. P. (1991). *An alternative approach to helical diffraction. Acta Cryst.* A**47**, 449–451.

Millane, R. P. (1992*a*). *Largest likely R factors for normal distributions. Acta Cryst.* A**48**, 649–650.

Millane, R. P. (1992*b*). *R factors in X-ray fibre diffraction. IV. Analytic expressions for largest likely R factors. Acta Cryst.* A**48**, 209–215.

Millane, R. P. (1993). *Image reconstruction from cylindrically averaged diffraction intensities.* In *Digital Image Recovery and Synthesis II*, Proc. SPIE, Vol. 2029, edited by P. S. Idell, pp. 137–143. Bellingham: SPIE.

Millane, R. P. & Arnott, S. (1985). *Background removal in X-ray fibre diffraction patterns. J. Appl. Cryst.* **18**, 419–423.

Millane, R. P. & Arnott, S. (1986). *Digital processing of X-ray diffraction patterns from oriented fibres. J. Macromol. Sci. Phys.* B**24**, 193–227.

Millane, R. P. & Baskaran, S. (1997). *Optimal difference Fourier synthesis in fibre diffraction. Fiber Diffr. Rev.* **6**, 14–18.

Millane, R. P., Byler, M. A. & Arnott, S. (1985). *Implementing constrained least squares refinement of helical polymers on a vector pipeline machine.* In *Supercomputer Applications*, edited by R. W. Numrich, pp. 137–143, New York: Plenum.

Millane, R. P., Chandrasekaran, R., Arnott, S. & Dea, I. C. M. (1988). *The molecular structure of kappa-carrageenan and comparison with iota-carrageenan. Carbohydr. Res.* **182**, 1–17.

Millane, R. P. & Stroud, W. J. (1991). *Effects of disorder on fibre diffraction patterns. Int. J. Biol. Macromol.* **13**, 202–208.

Millane, R. P. & Stubbs, G. (1992). *The significance of R factors in fibre diffraction. Polym. Prepr.* **33**, 321–322.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *On the application of the minimal principle to solve unknown structures. Science*, **259**, 1430–1433.

Namba, K., Pattanayek, R. & Stubbs, G. J. (1989). *Visualization of protein–nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fibre diffraction. J. Mol. Biol.* **208**, 307–325.

Namba, K. & Stubbs, G. (1985). *Solving the phase problem in fibre diffraction. Application to tobacco mosaic virus at 3.6 Å resolution. Acta Cryst.* A**41**, 252–262.

Namba, K. & Stubbs, G. (1987*a*). *Difference Fourier syntheses in fibre diffraction. Acta Cryst.* A**43**, 533–539.

Namba, K. & Stubbs, G. (1987*b*). *Isomorphous replacement in fibre diffraction using limited numbers of heavy-atom derivatives. Acta Cryst.* A**43**, 64–69.

Namba, K., Wakabayashi, K. & Mitsui, T. (1980). *X-ray structure analysis of the thin filament of crab striated muscle in the rigor state. J. Mol. Biol.* **138**, 1–26.

Namba, K., Yamashita, I. & Vonderviszt, F. (1989). *Structure of the core and central channel of bacterial flagella. Nature (London)*, **342**, 648–654.

Nambudripad, R., Stark, W. & Makowski, L. (1991). *Neutron diffraction studies of the structure of filamentous bacteriophage Pf1. J. Mol. Biol.* **220**, 359–379.

Park, H., Arnott, S., Chandrasekaran, R., Millane, R. P. & Campagnari, F. (1987). *Structure of the α-form of poly(dA)·poly(dT) and related polynucleotide duplexes. J. Mol. Biol.* **197**, 513–523.

Perez, S. & Chanzy, H. (1989). *Electron crystallography of linear polysaccharides. J. Electron Microsc. Tech.* **11**, 280–285.

Rickert, S. E., Lando, J. B., Hopfinger, A. J. & Baer, E. (1979). *Epitaxial polymerization of $(SN)_x$ 1. Structure and morphology of single crystals on alkali halide substrates. Macromolecules*, **12**, 1053–1057.

Rybnikar, F., Liu, J. & Geil, P. H. (1994). *Thin film melt-polymerized single crystals of poly(p-oxybenzoate). Makromol. Chem. Phys.* **195**, 81–104.

Sayre, D. (1952). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Schneider, A. I., Blackwell, J., Pielartzik, H. & Karbach, A. (1991). *Structure analysis of copoly(ester carbonate). Macromolecules*, **24**, 5676–5682.

Shotton, M. W., Denny, R. C. & Forsyth, V. T. (1998). *CCP13 software development. Fiber Diffr. Rev.* **7**, 40–44.

Sim, G. A. (1960). *A note on the heavy atom method. Acta Cryst.* **13**, 511–512.

Squire, J. M., Al-Khayat, H. A. & Yagi, N. (1993). *Muscle thin-filament structure and regulation. Actin sub-domain movements and the tropomyosin shift modelled from low-angle X-ray diffraction. J. Chem. Soc. Faraday Trans.* **89**, 2717–2726.

Squire, J., Cantino, M., Chew, M., Denny, R., Harford, J., Hudson, L. & Luther, P. (1998). *Myosin rod-packing schemes in vertebrate muscle thick filaments. J. Struct. Biol.* **122**, 128–138.

Squire, J. M. & Vibert, P. J. (1987). Editors. *Fibrous Protein Structure.* London: Academic Press.

Stanley, E. (1986). *'Peakiness' test functions. Acta Cryst.* A**42**, 297–299.

Stark, W., Glucksman, M. J. & Makowski, L. (1988). *Conformation of the coat protein of filamentous bacteriophage Pf1 determined by neutron diffraction from magnetically oriented gels of specifically deuterated virions. J. Mol. Biol.* **199**, 171–182.

Storks, K. H. (1938). *An electron diffraction examination of some linear high polymers. J. Am. Chem. Soc.* **60**, 1753–1761.

Stroud, W. J. & Millane, R. P. (1995*a*). *Analysis of disorder in biopolymer fibres. Acta Cryst.* A**51**, 790–800.

Stroud, W. J. & Millane, R. P. (1995*b*). *Diffraction by disordered polycrystalline fibres. Acta Cryst.* A**51**, 771–790.

Stroud, W. J. & Millane, R. P. (1996*a*). *Cylindrically averaged diffraction by distorted lattices. Proc. R. Soc. London*, **452**, 151–173.

Stroud, W. J. & Millane, R. P. (1996*b*). *Diffraction by polycrystalline fibres with correlated disorder. Acta Cryst.* A**52**, 812–829.

Stubbs, G. (1987). *The Patterson function in fibre diffraction*. In *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 548–557. Oxford University Press.

Stubbs, G. (1989). *The probability distributions of X-ray intensities in fibre diffraction: largest likely values for fibre diffraction R factors. Acta Cryst.* A**45**, 254–258.

Stubbs, G. (1999). *Developments in fiber diffraction. Curr. Opin. Struct. Biol.* **9**, 615–619.

Stubbs, G., Warren, S. & Holmes, K. (1977). *Structure of RNA and RNA binding site in tobacco mosaic virus from a 4 Å map calculated from X-ray fibre diagrams. Nature (London)*, **267**, 216–221.

Stubbs, G. J. (1974). *The effect of disorientation on the intensity distribution of non-crystalline fibres. II. Applications. Acta Cryst.* A**30**, 639–645.

Stubbs, G. J. & Diamond, R. (1975). *The phase problem for cylindrically averaged diffraction patterns. Solution by isomorphous replacement and application to tobacco mosaic virus. Acta Cryst.* A**31**, 709–718.

Stubbs, G. J. & Makowski, L. (1982). *Coordinated use of isomorphous replacement and layer-line splitting in the phasing of fibre diffraction data. Acta Cryst.* A**38**, 417–425.

Tadokoro, H. (1979). *Structure of crystalline polymers*. New York: Wiley.

Tanaka, S. & Naya, S. (1969). *Theory of X-ray scattering by disordered polymer crystals. J. Phys. Soc. Jpn*, **26**, 982–993.

Tatarinova, L. I. & Vainshtein, B. K. (1962). *Issledovanie poli-γ-metil-*L*-glutamata v α-forme metodom difraktsi elektronov. (Electron diffraction study of poly-γ-methyl-*L*-glutamate in the α-form.) Visokomol. Soed.* **4**, 261–269.

Tirion, M., ben Avraham, D., Lorenz, M. & Holmes, K. C. (1995). *Normal modes as refinement parameters for the F-actin model. Biophys. J.* **68**, 5–12.

Tsuji, M. (1989). *Electron microscopy*. In *Comprehensive Polymer Science*, Vol. 1. *Polymer Characterization*, edited by G. A. Allen, pp. 785–866. Oxford: Pergamon Press.

Vainshtein, B. K. (1964). *Structure analysis by electron diffraction*. Oxford: Pergamon Press.

Vainshtein, B. K. (1966). *Diffraction of X-rays by chain molecules*. Amsterdam: Elsevier.

Vainshtein, B. K. & Tatarinova, L. I. (1967). *The β-form of poly-γ-methyl-*L*-glutamate. Sov. Phys. Crystallogr.* **11**, 494–498.

Vibert, P. J. (1987). *Fibre diffraction methods*. In *Fibrous Protein Structure*, edited by J. M. Squire & P. J. Vibert, pp. 23–45. New York: Academic Press.

Wang, H., Culver, J. N. & Stubbs, G. (1997). *Structure of ribgrass mosaic virus at 2.9 Å resolution: evolution and taxonomy of tobamoviruses. J. Mol. Biol.* **269**, 769–779.

Wang, H. & Stubbs, G. (1993). *Molecular dynamics refinement against fibre diffraction data. Acta Cryst.* A**49**, 504–513.

Wang, H. & Stubbs, G. J. (1994). *Structure determination of cucumber green mottle mosaic virus by X-ray fibre diffraction. Significance for the evolution of tobamoviruses. J. Mol. Biol.* **239**, 371–384.

Welberry, T. R., Miller, G. H. & Carroll, C. E. (1980). *Paracrystals and growth-disorder models. Acta Cryst.* A**36**, 921–929.

Welsh, L. C., Symmons, M. F. & Marvin, D. A. (2000). *The molecular structure and structural transition of the α-helical capsid in filamentous bacteriophage Pf1. Acta Cryst.* D**56**, 137–150.

Welsh, L. C., Symmons, M. F., Sturtevant, J. M., Marvin, D. A. & Perham, R. N. (1998). *Structure of the capsid of Pf3 filamentous phage determined from X-ray fiber diffraction data at 3.1 Å resolution. J. Mol. Biol.* **283**, 155–177.

Wilson, A. J. C. (1950). *Largest likely values for the reliability index. Acta Cryst.* **3**, 397–399.

Wittmann, J. C. & Lotz, B. (1985). *Polymer decoration: the orientation of polymer folds as revealed by the crystallization of polymer vapor. J. Polym. Sci. Polym. Phys. Ed.* **23**, 205–226.

Wittmann, J. C. & Lotz, B. (1990). *Epitaxial crystallization of polymers on organic and polymeric substrates. Prog. Polym. Sci.* **15**, 909–948.

Wunderlich, B. (1973). *Macromolecular Physics*, Vol. 1. *Crystal Structure, Morphology, Defects*. New York: Academic Press.

Yamashita, I., Hasegawa, K., Suzuki, H., Vonderviszt, F., Mimori-Kiyosue, Y. & Namba, K. (1998). *Structure and switching of bacterial flagellar filaments studied by X-ray fiber diffraction. Nature Struct. Biol.* **5**, 125–132.

Zugenmaier, P. & Sarko, A. (1980). *The variable virtual bond*. In *Fibre Diffraction Methods*, ACS Symposium Series Vol. 141, edited by A. D. French & K. H. Gardner, pp. 225–237. Washington: American Chemical Society.

International Tables for Crystallography (2010). Vol. B, Chapter 4.6, pp. 590–624.

4.6. Reciprocal-space images of aperiodic crystals

By W. Steurer‡ and T. Haibach

4.6.1. Introduction

The discovery of materials with icosahedral diffraction symmetry (Shechtman *et al.*, 1984) was the main reason for the reassessment of the definition of *crystallinity* and for the introduction of the concept of *aperiodic crystals*. The first aperiodic crystal, *i.e.* a material with Bragg reflections not located only at reciprocal-lattice nodes, was identified long before (Dehlinger, 1927). In the following decades a wealth of incommensurately modulated phases and composite crystals were discovered. Nevertheless, only a few attempts have been made to develop a crystallography of aperiodic crystals; the most powerful of these was the higher-dimensional approach (see de Wolff, 1974, 1977; Janner & Janssen, 1979, 1980*a*,*b*; de Wolff *et al.*, 1981). In fact, incommensurate structures can be easily described using the higher-dimensional approach and also, fully equivalently, in a dual way: as a three-dimensional (3D) combination of one or more periodic basic structures and one or several modulation waves (de Wolff, 1984). However, with the discovery of quasicrystals and their noncrystalline symmetries, the latter approach failed and geometrical crystallography including the higher-dimensional approach received new attention. For more recent reviews of the crystallography of all three types of aperiodic crystals see van Smaalen (1995), of incommensurately modulated structures see Cummins (1990), of quasicrystals see Steurer (1990, 1996), of quasicrystals and their crystalline approximants see Goldman & Kelton (1993) and Kelton (1995). Textbooks on quasicrystals have been written by Janot (1994) and Senechal (1995).

According to the traditional crystallographic definition, an *ideal crystal* corresponds to an infinite 3D periodic arrangement of identical structure motifs. Its symmetry can be described by one of the 230 3D space groups. Mathematically, a periodic structure can be generated by the convolution of a function representing the structure motif with a lattice function. The structure motif can be given, for instance, by the electron-density distribution $\rho(\mathbf{r})$ of one primitive unit cell of the structure. The lattice function $g(\mathbf{r})$ is represented by a set of δ functions at the nodes $\mathbf{r} = \sum_{i=1}^{3} k_i \mathbf{a}_i$ of a 3D lattice Λ with basis $\mathbf{a}_i$, $i = 1, \ldots, 3$, and $k_i \in \mathbb{Z}$ ($\mathbb{Z}$ is the set of integer numbers). In reciprocal space, this convolution corresponds to the product of the Fourier transform $G(\mathbf{H})$ of the lattice function $g(\mathbf{r})$ and the Fourier transform $F(\mathbf{H}) = \int_v \rho(\mathbf{r}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r})\, \mathrm{d}\mathbf{r}$ of the structure motif $\rho(\mathbf{r})$. $G(\mathbf{H})$ is represented by the reciprocal lattice Λ^* decorated with δ functions on the reciprocal-lattice nodes $\mathbf{H} = \sum_{i=1}^{3} h_i \mathbf{a}_i^*$, with the reciprocal-basis vectors $\mathbf{a}_i^*$, $i = 1, \ldots, 3$, defined by $\mathbf{a}_i \cdot \mathbf{a}_j^* = \delta_{ij}$ and $h_i \in \mathbb{Z}$. The product $G(\mathbf{H}) \times F(\mathbf{H})$ is called the weighted reciprocal lattice; the weights are given by the structure factors $F(\mathbf{H})$. Thus, the characteristic feature of an ideal crystal in direct and reciprocal space is the existence of a lattice. In direct space, this lattice is decorated with identical structure motifs preserving translational and point symmetry in the framework of *space-group symmetry*. In reciprocal space, only the point symmetry between structure factors is maintained. The *Fourier spectrum* (or *Fourier image*, *i.e.* the Fourier transform) of the electron-density distribution of an ideal crystal consists of a countably infinite set of discrete Bragg peaks with a strictly defined minimum distance.

This crystal definition can be generalized to $n > 3$ dimensions. A d-dimensional (dD) *ideal aperiodic crystal* can be defined as a dD irrational section of an n-dimensional (nD, $n > d$) crystal with nD lattice symmetry. The intersection of the nD *hypercrystal* with the dD physical space is equivalent to a projection of the weighted nD reciprocal lattice $\Sigma^* = \left\{\mathbf{H} = \sum_{i=1}^{n} h_i \mathbf{d}_i^* | h_i \in \mathbb{Z}\right\}$ onto the dD physical space. The resulting set (Fourier module) $M^* = \left\{\mathbf{H}^{\parallel} = \sum_{i=1}^{n} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\right\}$ is countably dense. *Countably dense* means that the dense set of Bragg peaks can be mapped one-to-one onto the set of natural numbers. Hence, the Bragg reflections can be indexed with integer indices on an appropriate basis. The Fourier module of the projected reciprocal-lattice vectors $\mathbf{H}^{\parallel}$ has the structure of a $\mathbb{Z}$ module of rank n. A $\mathbb{Z}$ module is a free Abelian group, its rank n is given by the number of free generators (rationally independent vectors). The dimension of a $\mathbb{Z}$ module is that of the vector space spanned by it. The vectors $\mathbf{a}_i^*$ are the images of the vectors $\mathbf{d}_i^*$ projected onto the physical space $\mathbf{V}^{\parallel}$. Thus, by definition, the 3D reciprocal space of an ideal *aperiodic crystal* consists of a countably dense set of Bragg reflections only. Contrary to an ideal *crystal*, a minimum distance between Bragg reflections does not exist in an aperiodic one. In summary, it may be stressed that the terms *aperiodic* and *periodic* refer to properties of crystal structures in dD space. In nD space, as considered here, lattice symmetry is always present and, therefore, the term *crystal* is used.

Besides the *aperiodic crystals* mentioned above, other classes of *aperiodic structures* with strictly defined construction rules exist (see Axel & Gratias, 1995). Contrary to the kind of aperiodic crystals dealt with in this chapter, the Fourier spectra of aperiodic structures considered in the latter reference are continuous and contain only in a few cases additional sharp Bragg reflections (δ peaks).

Experimentally, the borderline between aperiodic crystals and their periodic approximations (*crystalline approximants*) is not sharply defined. Finite crystal size, static and dynamic disorder, chemical impurities and defects broaden Bragg peaks and cause diffuse diffraction phenomena. Furthermore, the resolution function of the diffraction equipment is limited.

However, the concept of describing an aperiodic structure as a dD physical-space section of an nD crystal (see Section 4.6.2) is only useful if it significantly simplifies the description of its structural order. Thus, depending on the shape of the *atomic surfaces*, which gives information on the atomic ordering, incommensurately modulated structures (IMSs, Sections 4.6.2.2 and 4.6.3.1), composite structures (CSs, Sections 4.6.2.3 and 4.6.3.2), or quasiperiodic structures (QSs, Sections 4.6.2.4 and 4.6.3.3) can be obtained from irrational cuts. The atomic surfaces are continuous $(n-d)$-dimensional objects for IMSs and CSs, and discrete $(n-d)$-dimensional objects for QSs. A class of aperiodic crystals with discrete fractal atomic surfaces also exists (Section 4.6.2.5). In this case the Hausdorff dimension (Hausdorff, 1919) of the atomic surface is not an integer number and smaller than $n-d$. The most outstanding characteristic feature of a fractal is its scale invariance: the object appears similar to itself 'from near as from far, that is, whatever the scale' (Gouyet, 1996).

To overcome the problems connected with experimental resolution, the translational symmetry of periodic crystals is used as a hard constraint in the course of the determination of their structures. Hence, space-group symmetry is taken for granted and only the local atomic configuration in a unit cell (actually, asymmetric unit) remains to be determined. In reciprocal space, this assumption corresponds to a condensation of Bragg reflections with finite full width at half maximum (FWHM) to δ peaks

‡ To whom correspondence should be addressed.

Copyright © 2010 International Union of Crystallography

accurately located at the reciprocal-lattice nodes. Diffuse diffraction phenomena are mostly neglected. This extrapolation to the existence of an ideal crystal is generally out of the question even if samples of very poor quality (high mosaicity, microdomain structure, defects, ...) are investigated.

The same practice is convenient for the determination of real aperiodic structures once the type of idealized aperiodic ordering is 'known'. Again, the global ordering principle is taken as a hard constraint. For instance, the question of whether a structure is commensurately or incommensurately modulated can only be answered within a given experimental resolution. Experimentally, the ratio of the wavelength of a modulation to the period of the underlying lattice can always be determined as a rational number only. Saying that a structure is incommensurately modulated, with the above ratio being an irrational number, simply means that the experimental results can be better understood, modelled and interpreted assuming an incommensurate modulation. For example, an incommensurate charge-density wave can be moved through an ideal crystal without changing the energy of the crystal. This is not so for a commensurate modulation. In some cases, the modulation period changes with temperature in discrete steps ('devil's staircase'), generating a series of commensurate superstructures ('lock-in structures'); in other cases, a continuous variation can be observed within the experimental resolution. The latter case will be described best by an incommensurately modulated structure.

However, if only the local structure of an aperiodic crystal is of interest, a structure analysis does not take much more experimental effort than for a regular crystal. In contrast, for the analysis of the global structure, *i.e.* the characterization of the type of its 'aperiodicity', diffraction experiments with the highest possible resolution are essential. Some problems connected with the structure analysis of aperiodic crystals are dealt with in Section 4.6.4.

To determine the long-range order – whether a real 'quasicrystal' is perfectly quasiperiodic, on average quasiperiodic, a crystalline approximant or a nanodomain structure – requires information from experiments that are sensitive to changes of the global structure. Hence, one needs diffraction experiments that allow the accurate determination of the spatial intensity distribution. Consequently, the limiting factors for such experiments are the maximum spatial and intensity resolution of the diffraction and detection equipment, as well as the size and quality of the sample. Nevertheless, the resolution available on state-of-the-art standard synchrotron-beamline equipment is sufficient to test whether the ordering of atoms in an aperiodic crystal reaches the same degree of perfection as found in high-quality silicon. Of course, the higher the sample quality the more necessary it is to account for dynamical diffraction effects such as reflection broadening and displacement. Otherwise, a misinterpretation may bias the global structure modelling.

The following sections present an aid to the characterization of aperiodic crystals based on information from diffraction experiments and give a survey of aperiodic crystals from the viewpoint of the experimentally accessible reciprocal space. Characteristic features of the diffraction patterns of the different types of aperiodic crystals are shown. A standard way of determining the metrics and finding the optimum nD embedding is described. Structure-factor formulae for general and special cases are given.

4.6.2. The n-dimensional description of aperiodic crystals

4.6.2.1. Basic concepts

An incommensurate modulation of a lattice-periodic structure destroys its translational symmetry in direct and reciprocal space. In the early seventies, a method was suggested by de Wolff (1974) for restoring the lost lattice symmetry by considering the diffraction pattern of an *incommensurately modulated structure* (IMS) as a projection of an nD reciprocal lattice upon the physical space. n, the dimension of the superspace, is always larger than or equal to d, the dimension of the physical space. This leads to a simple method for the description and characterization of IMSs as well as a variety of new possibilities in their structure analysis. The nD *embedding method* is well established today and can be applied to all aperiodic crystals with reciprocal-space structure equivalent to a $\mathbb{Z}$ module with finite rank n (Janssen, 1988). The dimension of the embedding space is determined by the rank of the $\mathbb{Z}$ module, *i.e.* by the number of reciprocal-basis vectors necessary to allow for indexing all Bragg reflections with integer numbers. The point symmetry of the 3D reciprocal space (Fourier spectrum) constrains the point symmetry of the nD reciprocal lattice and restricts the number of possible nD symmetry groups.

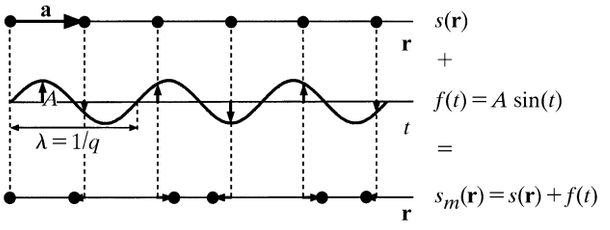


Fig. 4.6.2.1. The combination of a basic structure $s(\mathbf{r})$, with period a, and a sinusoidal modulation function $f(t)$, with amplitude A, period λ and $t = \mathbf{q} \cdot \mathbf{r}$, gives a modulated structure (MS) $s_m(\mathbf{r})$. The MS is aperiodic if a and λ are on incommensurate length scales. The filled circles represent atoms.

In the following sections, the nD descriptions of the four main classes of aperiodic crystals are demonstrated on simple 1D examples of incommensurately modulated phases, composite crystals, quasicrystals and structures with fractally shaped atomic surfaces. The main emphasis is placed on quasicrystals that show scaling symmetry, a new and unusual property in structural crystallography. A detailed discussion of the different types of 3D aperiodic crystals follows in Section 4.6.3.

4.6.2.2. 1D incommensurately modulated structures

A periodic deviation of atomic parameters from a reference structure (*basic structure*, BS) is called a *modulated structure* (MS). In the case of mutual incommensurability of the basic structure and the modulation period, the structure is called incommensurately modulated. Otherwise, it is called commensurately modulated. The modulated atomic parameters may be one or several of the following:

(a) coordinates,
(b) occupancy factors,
(c) thermal displacement parameters,
(d) orientation of the magnetic moment.

An incommensurately modulated structure can be described in a dual way by its *basic structure* $s(\mathbf{r})$ and a *modulation function* $f(t)$. This allows the structure-factor formula to be calculated and a full symmetry characterization employing representation theory to be performed (de Wolff, 1984). A more general method is the nD description: it relates the dD aperiodic incommensurately modulated structure to a periodic structure in nD space. This simplifies the symmetry analysis and structure-factor calculation, and allows more powerful structure-determination techniques.

The nD embedding method is demonstrated in the following 1D example of a displacively modulated structure. A basic structure $s(\mathbf{r}) = s(\mathbf{r} + n\mathbf{a})$, with period a and $n \in \mathbb{Z}$, is modulated by a function $f(t) = f(\mathbf{q} \cdot \mathbf{r}) = f(\alpha r) = f[\alpha r + (n a/\alpha)]$, with the satellite vector $\mathbf{q} = \alpha \mathbf{a}^*$, period $\lambda = 1/q = a/\alpha$, and α a rational or irrational number yielding a commensurately or incommensurately modulated structure $s_m(\mathbf{r})$ (Fig. 4.6.2.1).

If the 1D IMS and its 1D modulation function are properly combined in a 2D parameter space $\mathbf{V} = (\mathbf{V}^{\parallel}, \mathbf{V}^{\perp})$, a 2D lattice-

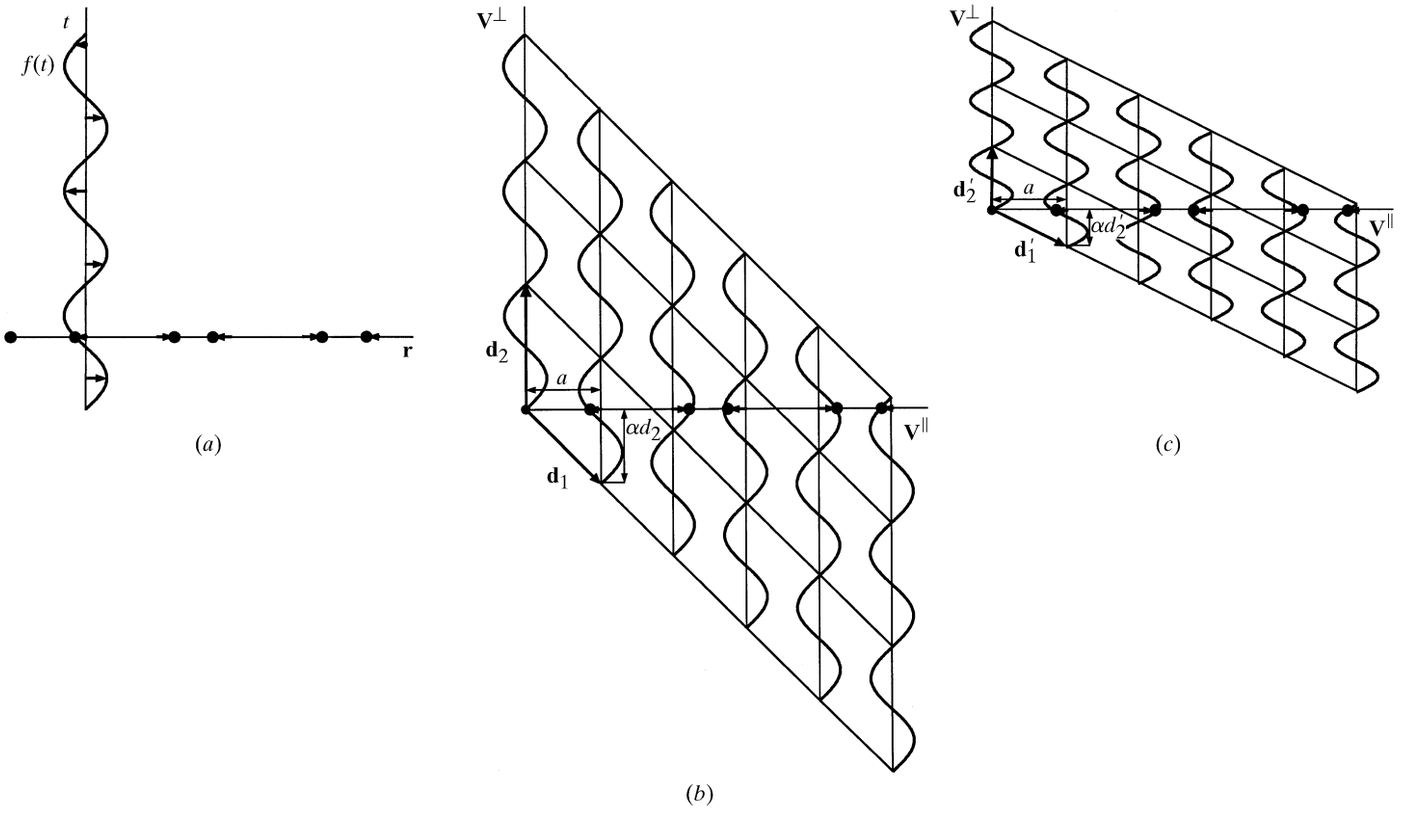


Fig. 4.6.2.2. 2D embedding of the sinusoidally modulated structure illustrated in Fig. 4.6.2.1. The correspondence between the actual displacement of an atom in the 1D structure and the modulation function defined in one additional dimension is illustrated in part (*a*). Adding to each atom its modulation function in this orthogonal dimension (perpendicular space $\mathbf{V}^{\perp}$) yields a periodic arrangement in 2D space $\mathbf{V}$, part (*b*). The MS results as a special section of the 2D periodic structure along the parallel space $\mathbf{V}^{\parallel}$. It is obvious from a comparison of (*b*) and (*c*) that the actual MS is independent of the perpendicular-space scale.

periodic structure results (Fig. 4.6.2.2). The actual atoms are generated by the intersection of the 1D physical (external, parallel) space $\mathbf{V}^{\parallel}$ with the continuous *hyperatoms*. The hyperatoms have the shape of the modulation function along the perpendicular (internal, complementary) space $\mathbf{V}^{\perp}$. They result from a convolution of the physical-space atoms with their modulation functions.

A basis $\mathbf{d}_1, \mathbf{d}_2$ (D basis) of the 2D hyperlattice $\Sigma = \{\mathbf{r} = \sum_{i=1}^{2} n_i \mathbf{d}_i | n_i \in \mathbb{Z}\}$ is given by

$$\mathbf{d}_1 = \begin{pmatrix} a \\ -\alpha/c \end{pmatrix}_V, \mathbf{d}_2 = \begin{pmatrix} 0 \\ 1/c \end{pmatrix}_V,$$

where a is the translation period of the BS and c is an arbitrary constant. The components of the basis vectors are given on a 2D orthogonal coordinate system (V basis). The components of the basis vector $\mathbf{d}_1$ are simply the parallel-space period a of the BS and α times the perpendicular-space component of the basis vector $\mathbf{d}_2$. The vector $\mathbf{d}_2$ is always parallel to the perpendicular space and its length is one period of the modulation function in arbitrary units (this is expressed by the arbitrary factor $1/c$). An atom at position $\mathbf{r}$ of the BS is displaced by an amount given by the modulation function $f(t)$, with $f(t) = f(\mathbf{q} \cdot \mathbf{r})$. Hence, the perpendicular-space variable t has to adopt the value $\mathbf{q} \cdot \mathbf{r} = \alpha \mathbf{a}^* \cdot r\mathbf{a} = \alpha r$ for the physical-space variable $\mathbf{r}$. This can be achieved by assigning the slope α to the basis vector $\mathbf{d}_1$. The choice of the parameter c has no influence on the actual MS, *i.e.* the way in which the 2D structure is cut by the parallel space (Fig. 4.6.2.2*c*).

The basis of the lattice $\Sigma^* = \{\mathbf{H} = \sum_{i=1}^{2} h_i \mathbf{d}_i^* | h_i \in \mathbb{Z}\}$, reciprocal to Σ, can be obtained from the condition $\mathbf{d}_i \cdot \mathbf{d}_j^* = \delta_{ij}$:

$$\mathbf{d}_1^* = \begin{pmatrix} a^* \\ 0 \end{pmatrix}_V, \mathbf{d}_2^* = \begin{pmatrix} \alpha a^* \\ c \end{pmatrix}_V,$$

with $a^* = 1/a$. The metric tensors for the reciprocal and direct 2D lattices for $c = 1$ are

$$G^* = \begin{pmatrix} a^{*2} & \alpha a^{*2} \\ \alpha a^{*2} & 1 + \alpha^2 a^{*2} \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} a^2 + \alpha^2 & -\alpha \\ -\alpha & 1 \end{pmatrix}.$$

The choice of an arbitrary number for c has no influence on the metrics of the physical-space components of the IMS in direct or reciprocal space.

The Fourier transform of the *hypercrystal* depicted in Fig. 4.6.2.2 gives the weighted reciprocal lattice shown in Fig. 4.6.2.3. The 1D diffraction pattern $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{2} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ in physical space is obtained by a projection of the weighted 2D reciprocal lattice Σ^* along $\mathbf{V}^{\perp}$ as the Fourier transform of a section in direct space corresponds to a projection in reciprocal space and *vice versa*:

$$M^* = \{\mathbf{H}^{\parallel}\} \xleftarrow{\text{projection } \pi^{\parallel} \text{ onto } \mathbf{V}^{\parallel}} \Sigma^* = \{\mathbf{H} = (\mathbf{H}^{\parallel}, \mathbf{H}^{\perp})\}.$$

Reciprocal-lattice points lying in physical space are referred to as *main reflections*, all others as *satellite reflections*. All Bragg reflections can be indexed with integer numbers h_1, h_2 in the 2D description $\mathbf{H} = h_1 \mathbf{d}_1^* + h_2 \mathbf{d}_2^*$. In the physical-space description, the diffraction vector can be written as $\mathbf{H}^{\parallel} = h\mathbf{a}^* + m\mathbf{q} = \mathbf{a}^*(h_1 + \alpha h_2)$, with $\mathbf{q} = \alpha \mathbf{a}^*$ for the satellite vector and $m \in \mathbb{Z}$ the order of the satellite reflection. For a detailed discussion of the

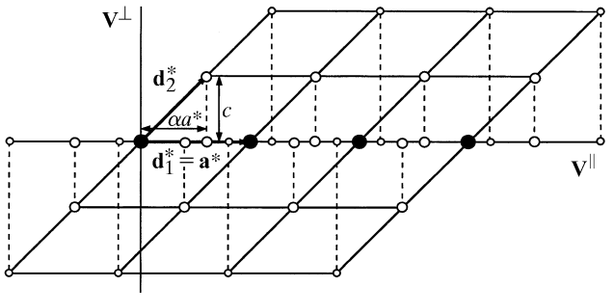


Fig. 4.6.2.3. Schematic representation of the 2D reciprocal-space embedding of the 1D sinusoidally modulated structure depicted in Figs. 4.6.2.1 and 4.6.2.2. Main reflections are marked by filled circles and satellite reflections by open circles. The sizes of the circles are roughly related to the reflection intensities. The actual 1D diffraction pattern of the 1D MS results from a projection of the 2D reciprocal space onto the parallel space. The correspondence between 2D reciprocal-lattice positions and their projected images is indicated by dashed lines.

embedding and symmetry description of IMSs see, for example, Janssen *et al.* (2004).

A commensurately modulated structure with $\alpha' = m/n$ and $\lambda = (n/m)a$, $m, n \in \mathbb{Z}$, and with $c = 1$, can be generated by shearing the 2D lattice Σ with a shear matrix S_m:

$$\mathbf{d}_i' = \sum_{j=1}^{2} S_{mij}\mathbf{d}_j, \text{ with } S_m = \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix}_D \text{ and } x = \alpha' - \alpha,$$

$$\mathbf{d}_1' = \mathbf{d}_1 - x\mathbf{d}_2 = \begin{pmatrix} a \\ -\alpha \end{pmatrix}_V - (\alpha' - \alpha)\begin{pmatrix} 0 \\ 1 \end{pmatrix}_V = \begin{pmatrix} a \\ -\alpha' \end{pmatrix}_V,$$

$$\mathbf{d}_2' = \mathbf{d}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_V.$$

The subscript D (V) following the shear matrix indicates that it is acting on the D (V) basis. The shear matrix does not change the distances between the atoms in the basic structure. In reciprocal space, using the inverted and transposed shear matrix, one obtains

$$\mathbf{d}_i^{*\prime} = \sum_{j=1}^{2} (S_m^{-1})_{ij}^T \mathbf{d}_j^*, \text{ with } (S_m^{-1})^T = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}_D \text{ and } x = \alpha' - \alpha,$$

$$\mathbf{d}_1^{*\prime} = \mathbf{d}_1^* = \begin{pmatrix} a^* \\ 0 \end{pmatrix}_V,$$

$$\mathbf{d}_2^{*\prime} = x\mathbf{d}_1^* + \mathbf{d}_2^* = (\alpha' - \alpha)\begin{pmatrix} a^* \\ 0 \end{pmatrix}_V + \begin{pmatrix} \alpha a^* \\ 1 \end{pmatrix}_V = \begin{pmatrix} \alpha' a^* \\ 1 \end{pmatrix}_V.$$

4.6.2.3. 1D composite structures

In the simplest case, a *composite structure* (CS) consists of two intergrown periodic structures with mutually incommensurate lattices. Owing to mutual interactions, each subsystem may be modulated with the period of the other. Consequently, CSs can be considered as coherent intergrowths of two or more incommensurately modulated substructures. The substructures have at least the origin of their reciprocal lattices in common. However, in all known cases, at least one common reciprocal-lattice plane exists. This means that at least one particular projection of the composite structure exhibits full lattice periodicity.

The unmodulated (basic) 1D subsystems of a 1D incommensurate intergrowth structure can be related to each other in a 2D parameter space $\mathbf{V} = (\mathbf{V}^{\parallel}, \mathbf{V}^{\perp})$ (Fig. 4.6.2.4). The actual atoms result from the intersection of the physical space $\mathbf{V}^{\parallel}$ with the hypercrystal. The hyperatoms correspond to a convolution of the

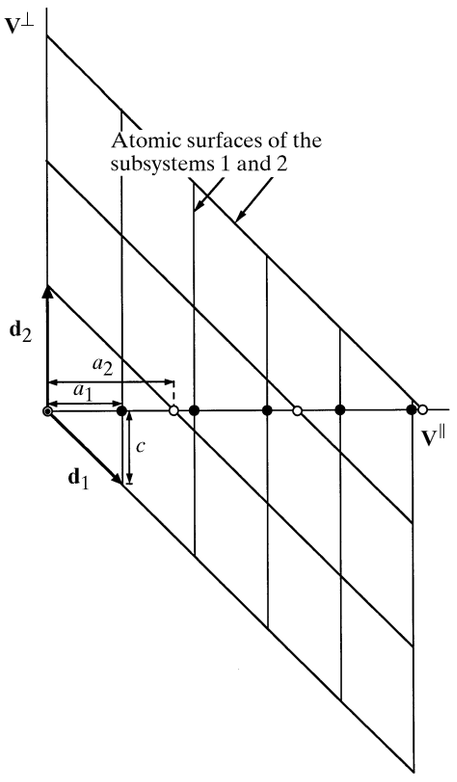


Fig. 4.6.2.4. 2D embedding of a 1D composite structure without mutual interaction of the subsystems. Filled and empty circles represent the atoms of the unmodulated substructures with periods a_1 and a_2, respectively. The atoms result from the parallel-space cut of the linear atomic surfaces parallel to $\mathbf{d}_1$ and $\mathbf{d}_2$.

real atoms with infinite lines parallel to the basis vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ of the 2D hyperlattice $\Sigma = \{\mathbf{r} = \sum_{i=1}^{2} n_i\mathbf{d}_i | n_i \in \mathbb{Z}\}$.

An appropriate basis is given by

$$\mathbf{d}_1 = \begin{pmatrix} a_1 \\ -c \end{pmatrix}_V, \mathbf{d}_2 = \begin{pmatrix} 0 \\ c(a_2/a_1) \end{pmatrix}_V,$$

where a_1 and a_2 are the lattice parameters of the two substructures and c is an arbitrary constant. Taking into account the interactions between the subsystems, each one becomes modulated with the period of the other. Consequently, in the 2D description, the shape of the hyperatoms is determined by their modulation functions (Fig. 4.6.2.5).

A basis of the reciprocal lattice $\Sigma^* = \{\mathbf{H} = \sum_{i=1}^{2} h_i\mathbf{d}_i^* | h_i \in \mathbb{Z}\}$ can be obtained from the condition $\mathbf{d}_i \cdot \mathbf{d}_j^* = \delta_{ij}$:

$$\mathbf{d}_1^* = \begin{pmatrix} a_1^* \\ 0 \end{pmatrix}_V, \mathbf{d}_2^* = \begin{pmatrix} a_2^* \\ (a_2^*/ca_1^*) \end{pmatrix}_V.$$

The metric tensors for the reciprocal and the direct 2D lattices for $c = 1$ are

$$G^* = \begin{pmatrix} a_1^{*2} & a_1^* a_2^* \\ a_1^* a_2^* & (1 + a_1^{*2})(a_2^*/a_1^*)^2 \end{pmatrix} \text{ and } G = \begin{pmatrix} 1 + a_1^2 & -a_2/a_1 \\ -a_2/a_1 & (a_2/a_1)^2 \end{pmatrix}.$$

The Fourier transforms of the hypercrystals depicted in Figs. 4.6.2.4 and 4.6.2.5 correspond to the weighted reciprocal lattices illustrated in Figs. 4.6.2.6 and 4.6.2.7. The 1D diffraction patterns $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{2} h_i\mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ in physical space are obtained by a projection of the weighted 2D reciprocal lattices Σ^* upon $\mathbf{V}^{\parallel}$. All Bragg reflections can be indexed with integer numbers h_1, h_2

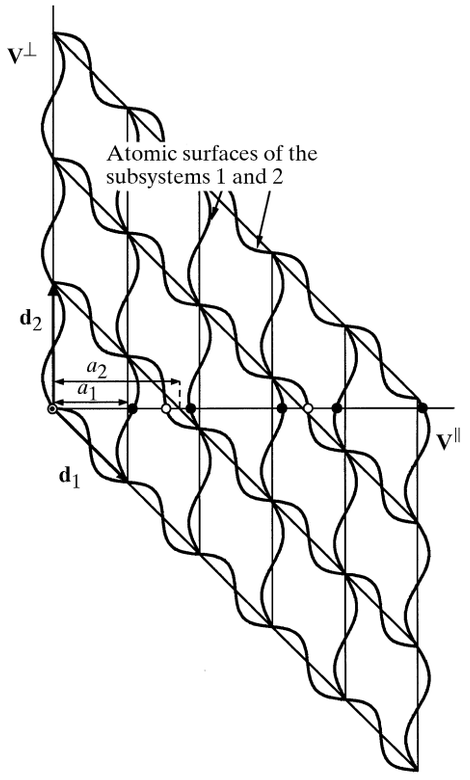


Fig. 4.6.2.5. 2D embedding of a 1D composite structure with mutual interaction of the subsystems causing modulations. Filled and empty circles represent the modulated substructures with periods a_1 and a_2 of the basic substructures, respectively. The atoms result from the parallel-space cut of the sinusoidal atomic surfaces running parallel to $\mathbf{d}_1$ and $\mathbf{d}_2$.

in both the 2D description $\mathbf{H} = h_1\mathbf{d}_1^* + h_2\mathbf{d}_2^*$ and in the 1D physical-space description with two parallel basis vectors $\mathbf{H}^{\parallel} = h_1\mathbf{a}_1^* + h_2\mathbf{a}_2^*$.

The reciprocal-lattice points $\mathbf{H} = h_1\mathbf{d}_1^*$ and $\mathbf{H} = h_2\mathbf{d}_2^*$, h_1, $h_2 \in \mathbb{Z}$, on the main axes $\mathbf{d}_1^*$ and $\mathbf{d}_2^*$ are the main reflections of the two substructures. All other reflections are referred to as satellite reflections. Their intensities differ from zero only in the case of modulated substructures. Each reflection of one subsystem coincides with exactly one reflection of the other subsystem.

4.6.2.4. 1D quasiperiodic structures

The Fibonacci sequence, the best investigated example of a 1D quasiperiodic structure, can be obtained from the substitution rule σ: S $\rightarrow$ L, L $\rightarrow$ LS, replacing the letter S by L and the letter L by the word LS (see *e.g.* Luck *et al.*, 1993). Applying the substitution matrix

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

associated with σ, this rule can be written in the form

$$\begin{pmatrix} \mathrm{S} \\ \mathrm{L} \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \mathrm{S} \\ \mathrm{L} \end{pmatrix} = \begin{pmatrix} \mathrm{L} \\ \mathrm{L}+\mathrm{S} \end{pmatrix}.$$

S gives the sum of letters, L + S = S + L, and not their order. Consequently, the same substitution matrix can also be applied, for instance, to the substitution σ': S $\rightarrow$ L, L $\rightarrow$ SL. The repeated action of S on the alphabet $A = \{\mathrm{S}, \mathrm{L}\}$ yields the words $A_n = \sigma^n(\mathrm{S})$ and $B_n = \sigma^n(\mathrm{L}) = A_{n+1}$ as illustrated in Table 4.6.2.1. The frequencies of letters contained in the words A_n and B_n can

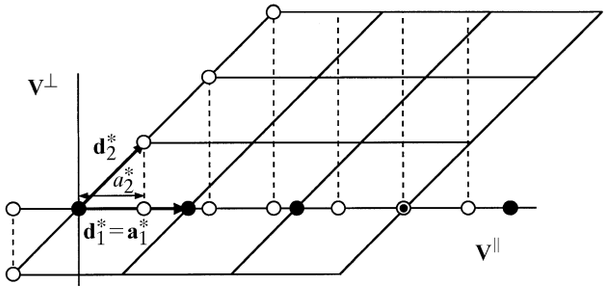


Fig. 4.6.2.6. Schematic representation of the reciprocal space of the embedded 1D composite structure depicted in Fig. 4.6.2.4. Filled and empty circles represent the reflections generated by the substructures with periods a_1 and a_2, respectively. The actual 1D diffraction pattern of the 1D CS results from a projection of the 2D reciprocal space onto the parallel space. The correspondence between 2D reciprocal-lattice positions and their projected images is indicated by dashed lines.

be calculated by applying the nth power of the transposed substitution matrix on the unit vector. From

$$\begin{pmatrix} \nu_{n+1}^A \\ \nu_{n+1}^B \end{pmatrix} = S^T \begin{pmatrix} \nu_n^A \\ \nu_n^B \end{pmatrix}$$

it follows that

$$\begin{pmatrix} \nu_n^A \\ \nu_n^B \end{pmatrix} = (S^T)^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

In the case of the Fibonacci sequence, ν_n^B gives the total number of letters S and L, and ν_n^A the number of letters L.

An infinite Fibonacci sequence, *i.e.* a word B_n with $n \rightarrow \infty$, remains invariant under inflation (deflation). Inflation (deflation) means that the number of letters L, S increases (decreases) under the action of the (inverted) substitution matrix S. Inflation and deflation represent self-similarity (scaling) symmetry operations on the infinite Fibonacci sequence. A more detailed discussion of the scaling properties of the Fibonacci chain in direct and reciprocal space will be given later.

The Fibonacci numbers $F_n = F_{n-1} + F_{n-2}$ form a series with $\lim_{n\rightarrow\infty}(F_{n+1}/F_n) = \tau$ {the golden mean $\tau = [1 + (5)^{1/2}]/2 = 2\cos(\pi/5) = 1.618\ldots$}. The ratio of the frequencies of L and S

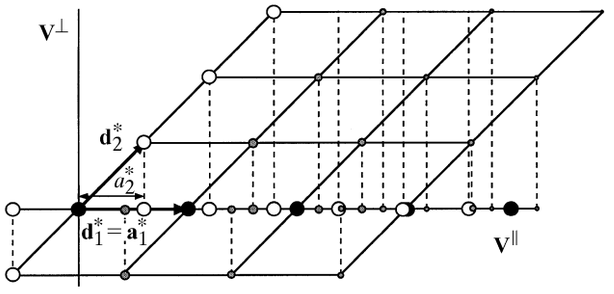


Fig. 4.6.2.7. Schematic representation of the reciprocal space of the embedded 1D composite structure depicted in Fig. 4.6.2.5. Filled and empty circles represent the main reflections of the two subsystems. The satellite reflections generated by the modulated substructures are shown as grey circles. The diameters of the circles are roughly proportional to the intensities of the reflections. The actual 1D diffraction pattern of the 1D CS results from a projection of the 2D reciprocal space onto the parallel space. The correspondence between 2D reciprocal-lattice positions and their projected images is indicated by dashed lines.

Table 4.6.2.1. *Expansion of the Fibonacci sequence* $B_n = \sigma^n(\mathrm{L})$ *by repeated action of the substitution rule* σ: $\mathrm{S} \rightarrow \mathrm{L}$, $\mathrm{L} \rightarrow \mathrm{LS}$

ν_L, ν_S are the frequencies of the letters L and S in word B_n.

B_n	ν_L	ν_S	n
L	1	0	0
LS	1	1	1
LSL	2	1	2
LSLLS	3	2	3
LSLLSLSL	5	3	4
LSLLSLSLLSLLS	8	5	5
LSLLSLSLLSLLSLSLLSLSL	13	8	6
	⋮	⋮	⋮
	F_{n+1}	F_n	n

in the Fibonacci sequence converges to τ if the sequence goes to infinity. The continued fraction expansion of the golden mean τ,

$$\tau = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ldots}}},$$

contains only the number 1. This means that τ is the 'most irrational' number, *i.e.* the irrational number with the worst truncated continued fraction approximation to it. This might be one of the reasons for the stability of quasiperiodic systems, where τ plays a role. The strong irrationality may impede the lock-in into commensurate systems (*rational approximants*).

By associating intervals (*e.g.* atomic distances) with length ratio τ to 1 to the letters L and S, a quasiperiodic structure $s(\mathbf{r})$ (*Fibonacci chain*) can be obtained. The invariance of the ratio of lengths $\mathrm{L/S} = (\mathrm{L} + \mathrm{S})/\mathrm{L} = \tau$ is responsible for the invariance of the Fibonacci chain under scaling by a factor τ^n, $n \in \mathbb{Z}$. Owing to a minimum atomic distance S in real crystal structures, the full set of inverse symmetry operators τ^{-n} does not exist. Consequently, the set of scaling operators $s = \{\tau^0 = 1, \tau^1, \ldots\}$ forms only a semi-group, *i.e.* an associative groupoid. Groupoids are the most general algebraic sets satisfying only one of the group axioms: the associative law. The scaling properties of the Fibonacci sequence can be derived from the eigenvalues of the scaling matrix S. For this purpose the equation

$$\det|S - \lambda I| = 0$$

with eigenvalue λ and unit matrix I has to be solved. The evaluation of the determinant yields the characteristic polynomial

$$\lambda^2 - \lambda - 1 = 0,$$

yielding in turn the eigenvalues $\lambda_1 = [1 + (5)^{1/2}]/2 = \tau$, $\lambda_2 = [1 - (5)^{1/2}]/2 = -1/\tau$ and the eigenvectors

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ \tau \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 1 \\ -1/\tau \end{pmatrix}.$$

Rewriting the eigenvalue equation gives for the first (*i.e.* the largest) eigenvalue

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ \tau \end{pmatrix} = \begin{pmatrix} \tau \\ 1 + \tau \end{pmatrix} = \begin{pmatrix} \tau \\ \tau^2 \end{pmatrix} = \tau\begin{pmatrix} 1 \\ \tau \end{pmatrix}.$$

Identifying the eigenvector

$$\begin{pmatrix} 1 \\ \tau \end{pmatrix}$$

with

$$\begin{pmatrix} \mathrm{S} \\ \mathrm{L} \end{pmatrix}$$

shows that an infinite Fibonacci sequence $s(\mathbf{r})$ remains invariant under scaling by a factor τ. This scaling operation maps each new lattice vector $\tau\mathbf{r}$ upon a vector $\mathbf{r}$ of the original lattice:

$$s(\tau\mathbf{r}) = s(\mathbf{r}).$$

Considering periodic lattices, these eigenvalues are integer numbers. For quasiperiodic 'lattices' (*quasilattices*) they always correspond to *algebraic numbers* (*Pisot numbers*). A Pisot number is the solution of a polynomial equation with integer coefficients. It is larger than one, whereas the modulus of its conjugate is smaller than unity: $\lambda_1 > 1$ and $|\lambda_2| < 1$ (Luck *et al.*, 1993). The total lengths l_n^A and l_n^B of the words A_n, B_n can be determined from the equations $l_n^A = \lambda_1^n l^A$ and $l_n^B = \lambda_1^n l^B$ with the eigenvalue λ_1. The left Perron–Frobenius eigenvector $\mathbf{w}_1$ of S, *i.e.* the left eigenvector associated with λ_1, determines the ratio S:L to 1:τ. The right Perron–Frobenius eigenvector $\mathbf{w}_1$ of S associated with λ_1 gives the relative frequencies, 1 and τ, for the letters S and L (for a definition of the Perron–Frobenius theorem see Luck *et al.*, 1993, and references therein).

The general case of an alphabet $A = \{\mathrm{L}_1 \ldots \mathrm{L}_k\}$ with k letters (intervals) L_i, of which at least two are on incommensurate length scales and which transform with the substitution matrix S,

$$\mathrm{L}_i' \rightarrow \sum_{j=1}^{k} S_{ij}\mathrm{L}_j,$$

can be treated analogously. S is a $k \times k$ matrix with non-negative integer coefficients. Its eigenvalues are solutions of a polynomial equation of rank k with integer coefficients (algebraic or Pisot numbers). The dimension n of the embedding space is generically equal to the number of letters (intervals) k involved by the substitution rule. From substitution rules, infinitely many different 1D quasiperiodic sequences can be generated. However, their atomic surfaces in the nD description are generically of fractal shape (see Section 4.6.2.5).

The quasiperiodic 1D density distribution $\rho(\mathbf{r})$ of the Fibonacci chain can be represented by the Fourier series

$$\rho(\mathbf{r}) = (1/V)\sum_{\mathbf{H}^\parallel} F(\mathbf{H}^\parallel)\exp(-2\pi i\mathbf{H}^\parallel \cdot \mathbf{r}),$$

with $\mathbf{H}^\parallel \in \mathbb{R}$ (the set of real numbers). The Fourier coefficients $F(\mathbf{H}^\parallel)$ form a Fourier module $M^* = \{\mathbf{H}^\parallel = \sum_{i=1}^{2} h_i\mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ equivalent to a $\mathbb{Z}$ module of rank 2. Thus a periodic function in 2D space can be defined by

$$\rho(\mathbf{r}^\parallel, \mathbf{r}^\perp) = (1/V)\sum_{\mathbf{H}} F(\mathbf{H})\exp[-2\pi i(\mathbf{H}^\parallel \cdot \mathbf{r}^\parallel + \mathbf{H}^\perp \cdot \mathbf{r}^\perp)],$$

where $\mathbf{r} = (\mathbf{r}^\parallel, \mathbf{r}^\perp) \in \Sigma$ and $\mathbf{H} = (\mathbf{H}^\parallel, \mathbf{H}^\perp) \in \Sigma^*$ are, by construction, direct- and reciprocal-lattice vectors (Figs. 4.6.2.8 and 4.6.2.9):

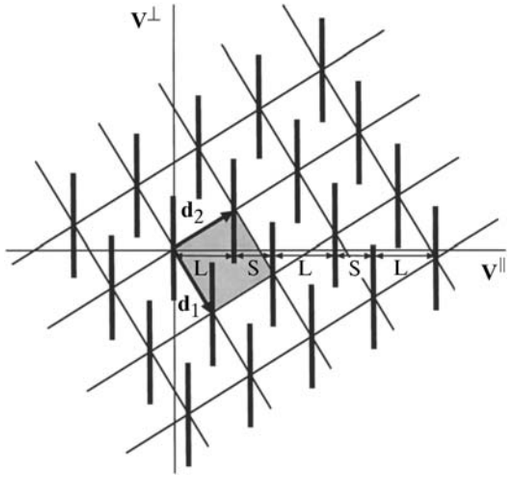


Fig. 4.6.2.8. 2D embedding of the Fibonacci chain. The short and long distances S and L, generated by the intersection of the atomic surfaces with the physical space $\mathbf{V}^\parallel$, are indicated. The atomic surfaces are represented by bars parallel to $\mathbf{V}^\perp$. Their lengths correspond to the projection of one unit cell (shaded) upon $\mathbf{V}^\perp$.

$$\mathbf{r} = n_1\mathbf{d}_1 + n_2\mathbf{d}_2, \text{ with } \mathbf{d}_1 = \frac{1}{a^*(2+\tau)}\begin{pmatrix}1\\-\tau\end{pmatrix}_V,$$

$$\mathbf{d}_2 = \frac{1}{a^*(2+\tau)}\begin{pmatrix}\tau\\1\end{pmatrix}_V;$$

$$\mathbf{H} = h_1\mathbf{d}_1^* + h_2\mathbf{d}_2^*, \text{ with } \mathbf{d}_1^* = a^*\begin{pmatrix}1\\-\tau\end{pmatrix}_V, \mathbf{d}_2^* = a^*\begin{pmatrix}\tau\\1\end{pmatrix}_V.$$

The 1D Fibonacci chain results from the cut of the parallel (physical) space with the 2D lattice Σ decorated with line elements for the *atomic surfaces* (*acceptance domains*). In this description, the atomic surfaces correspond simply to the projection of one 2D unit cell upon the perpendicular-space coordinate. This satisfies the condition that each unit cell contributes exactly to one point of the Fibonacci chain (*primitive unit cell*). The physical space $\mathbf{V}^\parallel$ is related to the eigenspace of the substitution matrix S associated with its eigenvalue $\lambda_1 = \tau$. The perpendicular space $\mathbf{V}^\perp$ corresponds to the eigenspace of the substitution matrix S associated with its eigenvalue $\lambda_2 = -1/\tau$. Thus, the physical space scales to powers of τ and the perpendicular space to powers of $-1/\tau$.

By block-diagonalization, the reducible substitution (scaling) matrix S can be decomposed into two non-equivalent irreducible representations. These can be assigned to the two 1D orthogonal subspaces $\mathbf{V}^\parallel$ and $\mathbf{V}^\perp$ forming the 2D embedding space $\mathbf{V} = \mathbf{V}^\parallel \oplus \mathbf{V}^\perp$. Thus, using $WSW^{-1} = S_V = S_V^\parallel \oplus S_V^\perp$, where

$$W = \begin{pmatrix}1 & \tau\\ -\tau & 1\end{pmatrix} = \begin{pmatrix}\mathbf{d}_1^* & \mathbf{d}_2^*\end{pmatrix},$$

one obtains

$$\begin{pmatrix}1 & \tau\\ -\tau & 1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}_D\begin{pmatrix}1 & \tau\\ -\tau & 1\end{pmatrix}^{-1} = \left(\begin{array}{c|c}\tau & 0\\ \hline 0 & -1/\tau\end{array}\right)_V = \left(\begin{array}{c|c}S^\parallel & 0\\ \hline 0 & S^\perp\end{array}\right)_V,$$

the scaling operations $S^\parallel$ and $S^\perp$ in parallel and in perpendicular space as indicated by the partition lines.

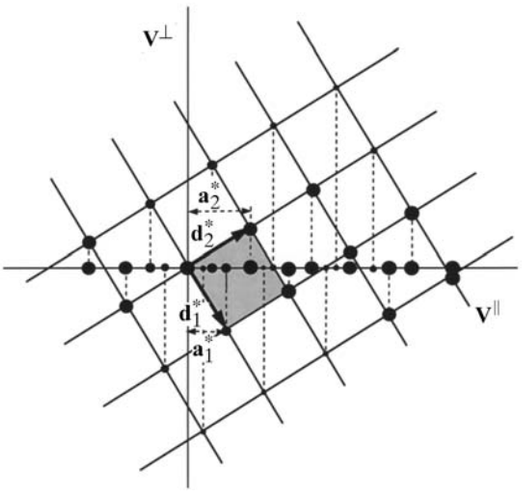


Fig. 4.6.2.9. Schematic representation of the reciprocal space of the embedded Fibonacci chain depicted in Fig. 4.6.2.8. The physical-space reciprocal basis $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$ is marked. The diameters of the filled circles are roughly proportional to the reflection intensities. One 2D reciprocal-lattice unit cell is shadowed. The actual 1D diffraction pattern of the 1D Fibonacci chain results from a projection of the 2D reciprocal space onto the parallel space. The correspondence between 2D reciprocal-lattice positions and their projected images is indicated by dashed lines.

The metric tensors for the reciprocal and the direct 2D square lattices read

$$G^* = |a^*|^2(2+\tau)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \text{ and } G = \frac{1}{|a^*|^2(2+\tau)}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

The short distance S of the Fibonacci sequence is related to a^* by

$$\begin{aligned} S &= 1/[a^*(2+\tau)]\\ &= \min\left\{\left|\pi^\parallel(\mathbf{d}_i - \mathbf{d}_j)\right| \,\middle|\, \left|\pi^\perp(\mathbf{d}_i - \mathbf{d}_j)\right| < \Omega_{\rm AS} \wedge i, j \in \mathbb{Z}\right\}, \end{aligned}$$

with the projectors $\pi^\parallel$ and $\pi^\perp$ onto $\mathbf{V}^\parallel$ and $\mathbf{V}^\perp$. The *point density* $\rho_{\rm p}$ of the Fibonacci chain, *i.e.* the number of vertices per unit length, can be calculated using the formula

$$\rho_{\rm p} = \frac{\Omega_{\rm AS}}{\Omega_{\rm UC}} = \frac{(1+\tau)/[a^*(2+\tau)]}{1/[|a^*|^2(2+\tau)]} = a^*\tau^2,$$

where $\Omega_{\rm AS}$ and $\Omega_{\rm UC}$ are the areas of the atomic surface and of the 2D unit cell, respectively.

For an infinite Fibonacci sequence generated from the intervals S and L an average distance d can be calculated:

$$d = \lim_{n\to\infty}\frac{F_n\mathrm{S} + F_{n+1}\mathrm{L}}{F_n + F_{n+1}} = \lim_{n\to\infty}\frac{F_n(\mathrm{S}+\tau\mathrm{L})}{F_n(1+\tau)} = \frac{\mathrm{S}(1+\tau^2)}{(1+\tau)} = (3-\tau)\mathrm{S}.$$

Therefrom, the point density can also be calculated:

$$\rho_{\rm p} = 1/d = 1/[(3-\tau)\mathrm{S}] = [a^*(2+\tau)]/(3-\tau) = a^*\tau^2.$$

An approximant structure of the Fibonacci sequence with a unit cell containing m intervals L and n intervals S can be generated by shearing the 2D lattice Σ by the shear matrix S_m,

$$S_m = \frac{1}{2+\tau}\begin{pmatrix} \tau^2 + x\tau + 1 & -x \\ x\tau^2 & \tau^2 - x\tau + 1 \end{pmatrix}_D,$$

where $x = (n\tau - m)/(m\tau + n)$:

$$\begin{aligned}
\mathbf{d}_i' &= \sum_{j=1}^{2} S_{mij}\mathbf{d}_j; \\
\mathbf{d}_1' &= \frac{1}{2+\tau}\left[(\tau^2 + x\tau + 1)\mathbf{d}_1 - x\mathbf{d}_2\right] \\
&= \frac{1}{(2+\tau)a^*}\begin{pmatrix} 1 \\ -\tau - x \end{pmatrix}_V \\
&= \frac{1}{(2+\tau)a^*}\begin{pmatrix} 1 \\ -\dfrac{2n\tau + m\tau}{m\tau + n} \end{pmatrix}_V, \\
\mathbf{d}_2' &= \frac{1}{2+\tau}\left[x\tau^2\mathbf{d}_1 + (\tau^2 - x\tau + 1)\mathbf{d}_2\right] \\
&= \frac{1}{(2+\tau)a^*}\begin{pmatrix} \tau \\ -x\tau + 1 \end{pmatrix}_V \\
&= \frac{1}{(2+\tau)a^*}\begin{pmatrix} \tau \\ \dfrac{2m\tau - n\tau}{m\tau + n} \end{pmatrix}_V.
\end{aligned}$$

This shear matrix does not change the magnitudes of the intervals L and S. In reciprocal space the inverted and transposed shear matrix is applied on the reciprocal basis,

$$(S_m^{-1})^T = \frac{1}{2+\tau}\begin{pmatrix} \tau^2 - x\tau + 1 & -x\tau^2 \\ x & \tau^2 + x\tau + 1 \end{pmatrix}_D,$$

where $x = (n\tau - m)/(m\tau + n)$:

$$\begin{aligned}
\mathbf{d}_i^{*'} &= \sum_{j=1}^{2} (S_m^{-1})_{ij}^T \mathbf{d}_j^*; \\
\mathbf{d}_1^{*'} &= \frac{1}{2+\tau}\left[(\tau^2 - x\tau + 1)\mathbf{d}_1^* - x\tau^2\mathbf{d}_2^*\right] \\
&= a^*\begin{pmatrix} 1 - x\tau \\ -\tau \end{pmatrix}_V \\
&= a^*\begin{pmatrix} \dfrac{2m\tau - n\tau}{m\tau + n} \\ -\tau \end{pmatrix}_V, \\
\mathbf{d}_2^{*'} &= \frac{1}{2+\tau}\left[x\mathbf{d}_1^* + (\tau^2 + x\tau + 1)\mathbf{d}_2^*\right] \\
&= a^*\begin{pmatrix} \tau + x \\ 1 \end{pmatrix}_V \\
&= a^*\begin{pmatrix} \dfrac{2n\tau + m\tau}{m\tau + n} \\ 1 \end{pmatrix}_V.
\end{aligned}$$

The point $x_n(t)$ of the nth interval L or S of an infinite Fibonacci sequence is given by

$$x_n(t) = \{x_0 + n(3 - \tau) - (\tau - 1)[\mathrm{frac}(n\tau + t) - (1/2)]\}\mathrm{S},$$

where t is the phase of the modulation function $y(t) = (\tau - 1)[\mathrm{frac}(n\tau + t) - (1/2)]$ (Janssen, 1986). Thus, the Fibonacci sequence can also be dealt with as an incommensurately modulated structure. This is a consequence of the fact that for 1D structures only the crystallographic point symmetries 1 and $\bar{1}$ allow the existence of a periodic average structure.

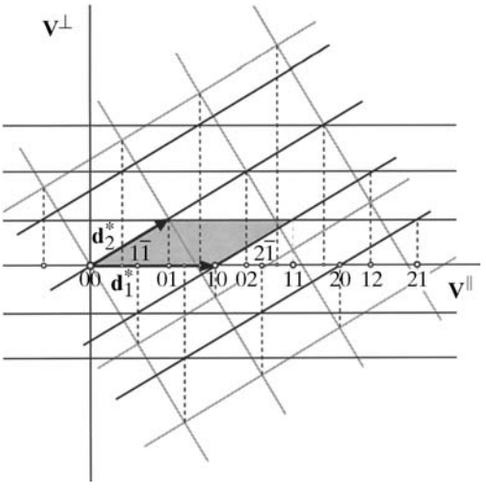


Fig. 4.6.2.10. Reciprocal space of the embedded Fibonacci chain as a modulated structure. Several main and satellite reflections are indexed. The square reciprocal lattice of the quasicrystal description illustrated in Fig. 4.6.2.9 is indicated by grey lines. The reflections located on $\mathbf{V}^\parallel$ can be considered to be projected either from the 2D square lattice of the embedding as for a QS or from the 2D oblique lattice of the embedding as for an IMS.

The embedding of the Fibonacci chain as an incommensurately modulated structure can be performed as follows:

(1) select a subset $\Lambda^* \subset M^*$ of strong reflections for main reflections $\mathbf{H} = h\mathbf{a}^*, h \in \mathbb{Z}$;

(2) define a satellite vector $\mathbf{q} = \alpha\mathbf{a}^*$ pointing from each main reflection to the next satellite reflection.

One possible way of indexing based on the same $\mathbf{a}^*$ as defined above is illustrated in Fig. 4.6.2.10. The scattering vector is given by $\mathbf{H}^\parallel = h(\tau + 1)\mathbf{a}^* + mq$, where $\mathbf{q} = \tau\mathbf{a}^*$, or, in the 2D representation, $\mathbf{H} = h_1\mathbf{d}_1^* + h_2\mathbf{d}_2^*$, where

$$\mathbf{d}_1^* = a^*\begin{pmatrix} 1 + \tau \\ 0 \end{pmatrix}_V$$

and

$$\mathbf{d}_2^* = a^*\begin{pmatrix} \tau \\ 1 \end{pmatrix}_V,$$

with the direct basis

$$\mathbf{d}_1 = \frac{1}{a^*(1+\tau)}\begin{pmatrix} 1 \\ -\tau \end{pmatrix}_V, \quad \mathbf{d}_2 = \frac{1}{a^*}\begin{pmatrix} 0 \\ 1 \end{pmatrix}_V.$$

The modulation function is saw-tooth-like (Fig. 4.6.2.11).

4.6.2.5. 1D structures with fractal atomic surfaces

A 1D structure with a *fractal atomic surface* (Hausdorff dimension 0.9157...) can be derived from the Fibonacci sequence by squaring its substitution matrix S:

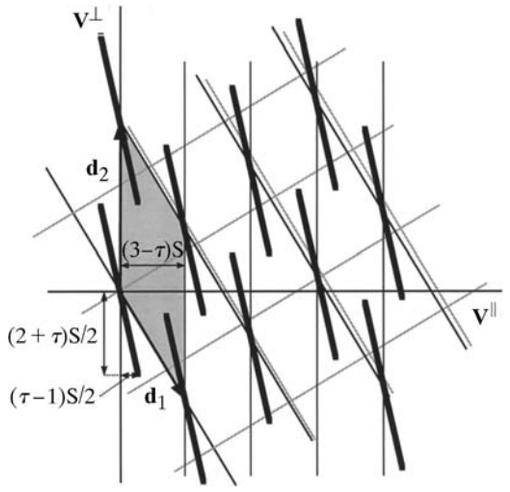


Fig. 4.6.2.11. 2D direct-space embedding of the Fibonacci chain as a modulated structure. The average period is $(3-\tau)$S. The square lattice in the quasicrystal description shown in Fig. 4.6.2.8 is indicated by grey lines. The rod-like atomic surfaces are now inclined relative to $\mathbf{V}^{\parallel}$ and arranged so as to give a saw-tooth modulation wave.

$$\begin{pmatrix} \mathrm{S} \\ \mathrm{L} \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \mathrm{S} \\ \mathrm{L} \end{pmatrix} = \begin{pmatrix} \mathrm{S}+\mathrm{L} \\ \mathrm{S}+2\mathrm{L} \end{pmatrix}$$
$$\text{with } S^2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

corresponding to the substitution rule S $\rightarrow$ SL, L $\rightarrow$ LLS as well as two other non-equivalent ones (see Janssen, 1995). The eigenvalues λ_i are obtained by calculating

$$\det |S - \lambda I| = 0.$$

The evaluation of the determinant gives the characteristic polynomial

$$\lambda^2 - 3\lambda + 1 = 0,$$

with the solutions $\lambda_{1,2} = [3 \pm (5)^{1/2}]/2$, with $\lambda_1 = \tau^2$ and $\lambda_2 = 1/\tau^2 = 2 - \tau$, and the same eigenvectors

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ \tau \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 1 \\ -1/\tau \end{pmatrix}$$

as for the Fibonacci sequence. Rewriting the eigenvalue equation gives

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ \tau \end{pmatrix} = \begin{pmatrix} \tau+1 \\ 2\tau+1 \end{pmatrix} = \begin{pmatrix} \tau^2 \\ \tau^3 \end{pmatrix} = \tau^2 \begin{pmatrix} 1 \\ \tau \end{pmatrix}.$$

Identifying the eigenvector

$$\begin{pmatrix} 1 \\ \tau \end{pmatrix}$$

with

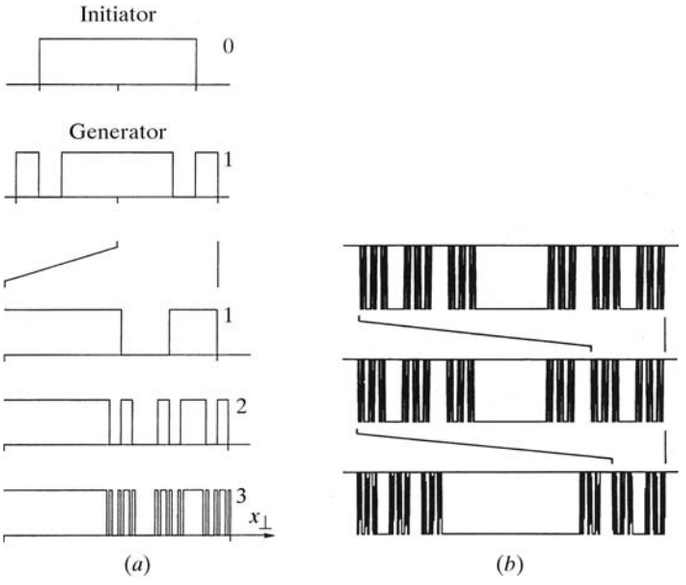


Fig. 4.6.2.12. (*a*) Three steps in the development of the fractal atomic surface of the squared Fibonacci sequence starting from an initiator and a generator. The action of the generator is to cut a piece from each side of the initiator and to add it where the initiator originally ended. This is repeated, cutting thinner and thinner pieces each time from the generated structures. (*b*) Magnification sequence of the fractal atomic surface illustrating its self-similarity. Each successive figure represents a magnification of a selected portion of the previous figure (from Zobetz, 1993).

$$\begin{pmatrix} \mathrm{S} \\ \mathrm{L} \end{pmatrix}$$

shows that the infinite 1D sequence $s(\mathbf{r})$ multiplied by powers of its eigenvalue τ^2 (scaling operation) remains invariant (each new lattice point coincides with one of the original lattice):

$$s(\tau^2\mathbf{r}) = s(\mathbf{r}).$$

The fractal sequence can be described on the same reciprocal and direct bases as the Fibonacci sequence. The only difference in the 2D direct-space description is the fractal character of the perpendicular-space component of the hyperatoms (Fig. 4.6.2.12) (see Zobetz, 1993).

4.6.3. Reciprocal-space images

4.6.3.1. Incommensurately modulated structures (IMSs)

One-dimensionally modulated structures are the simplest representatives of IMSs. The vast majority of the one hundred or so IMSs known so far belong to this class (Cummins, 1990). However, there is also an increasing number of IMSs with 2D or 3D modulation. The dimension d of the modulation is defined by the number of rationally independent modulation wavevectors (satellite vectors) $\mathbf{q}_i$ (Fig. 4.6.3.1). The electron-density function of a dD modulated 3D crystal can be represented by the Fourier series

$$\rho(\mathbf{r}) = (1/V)\sum_{\mathbf{H}} F(\mathbf{H}) \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}).$$

The Fourier coefficients (*structure factors*) $F(\mathbf{H})$ differ from zero only for reciprocal-space vectors $\mathbf{H} = \sum_{i=1}^{3} h_i \mathbf{a}_i^* + \sum_{j=1}^{d} m_j \mathbf{q}_j = \sum_{i=1}^{3+d} h_i \mathbf{a}_i^*$ with $h_i, m_j \in \mathbb{Z}$. The d satellite vectors are given by $\mathbf{q}_j = \mathbf{a}_{3+j}^* = \sum_{i=1}^{3} \alpha_{ij} \mathbf{a}_i^*$, with α_{ij} a $3 \times d$ matrix σ. In the case of an

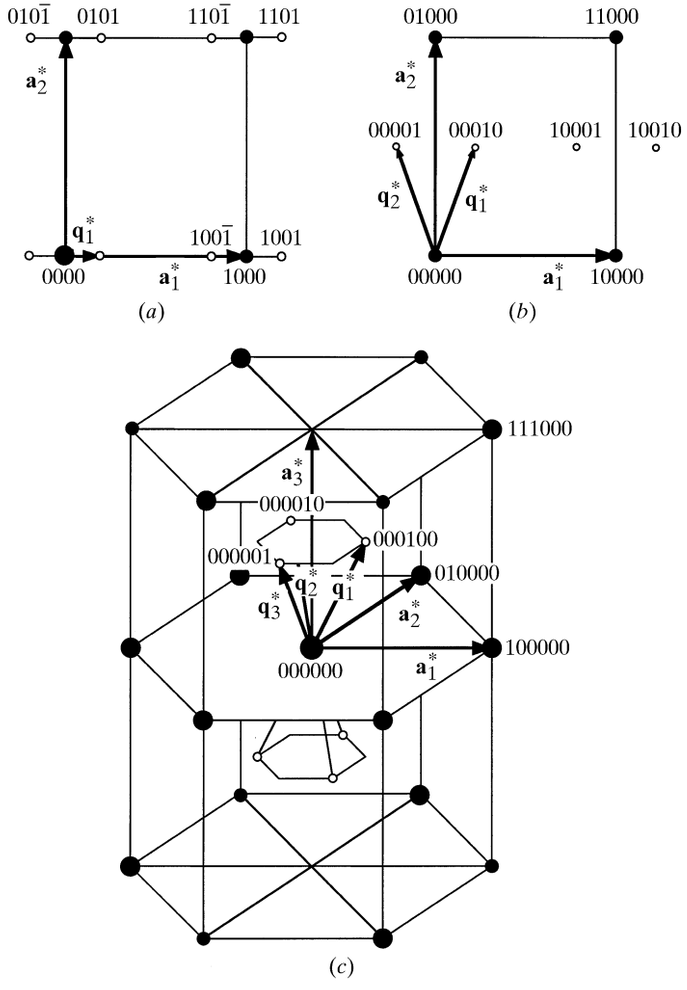


Fig. 4.6.3.1. Schematic diffraction patterns for IMSs with (*a*) 1D, (*b*) 2D and (*c*) 3D modulation. The satellite vectors correspond to $\mathbf{q} = \alpha_1\mathbf{a}_1^*$ in (*a*), $\mathbf{q}_1 = \alpha_{11}\mathbf{a}_1^* + (1/2)\mathbf{a}_2^*$ and $\mathbf{q}_2 = -\alpha_{12}\mathbf{a}_1^* + (1/2)\mathbf{a}_2^*$, where $\alpha_{11} = \alpha_{12}$, in (*b*), and $\mathbf{q}_1 = \alpha_{11}\mathbf{a}_1^* + \alpha_{31}\mathbf{a}_3^*$, $\mathbf{q}_2 = \alpha_{12}(-\mathbf{a}_1^* + \mathbf{a}_2^*) + \alpha_{32}\mathbf{a}_3^*$, $\mathbf{q}_3 = -\alpha_{13}\mathbf{a}_2^* + \alpha_{33}\mathbf{a}_3^*$, where $\alpha_{11} = \alpha_{12} = \alpha_{13}$ and $\alpha_{31} = \alpha_{32} = \alpha_{33}$, in (*c*). The areas of the circles are proportional to the reflection intensities. Main (filled circles) and satellite (open circles) reflections are indexed (after Janner *et al.*, 1983*b*).

IMS, at least one entry to σ has to be irrational. The wavelength of the modulation function is $\lambda_j = 1/q_j$. The set of vectors $\mathbf{H}$ forms a Fourier module $M^* = \{\mathbf{H} = \sum_{i=1}^{3+d} h_i\mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ of rank $n = 3 + d$, which can be decomposed into a rank 3 and a rank d submodule $M^* = M_1^* \oplus M_2^*$. $M_1^* = \{h_1\mathbf{a}_1^* + h_2\mathbf{a}_2^* + h_3\mathbf{a}_3^*\}$ corresponds to a $\mathbb{Z}$ module of rank 3 in a 3D subspace (the physical space), $M_2^* = \{h_4\mathbf{a}_4^* + \ldots + h_{3+d}\mathbf{a}_{3+d}^*\}$ corresponds to a $\mathbb{Z}$ module of rank d in a dD subspace (perpendicular space). The submodule M_1 is identical to the 3D reciprocal lattice Λ^* of the average structure. M_2 results from the projection of the perpendicular-space component of the $(3 + d)$D reciprocal lattice Σ^* upon the physical space. Owing to the coincidence of one subspace with the physical space, the dimension of the embedding space is given as $(3 + d)$D and not as nD. This terminology points out the special role of the physical space.

Hence the reciprocal-basis vectors $\mathbf{a}_i^*, i = 1, \ldots, 3 + d$, can be considered to be physical-space projections of reciprocal-basis vectors $\mathbf{d}_i^*, i = 1, \ldots, 3 + d$, spanning a $(3 + d)$D reciprocal lattice Σ^*:

$$\Sigma^* = \left\{\mathbf{H} = \sum_{i=1}^{3+d} h_i\mathbf{d}_i^* \Big| h_i \in \mathbb{Z}\right\},$$

$$\mathbf{d}_i^* = (\mathbf{a}_i^*, \mathbf{0}),\ i = 1, \ldots, 3 \text{ and } \mathbf{d}_{3+j}^* = (\mathbf{a}_{3+j}^*, c\mathbf{e}_j^*),\ j = 1, \ldots, d.$$

The first vector component of $\mathbf{d}_i^*$ refers to the physical space, the second to the perpendicular space spanned by the mutually orthogonal unit vectors $\mathbf{e}_j$. c is an arbitrary constant which can be set to 1 without loss of generality.

A direct lattice Σ with basis $\mathbf{d}_i$, $i = 1, \ldots, 3 + d$ and $\mathbf{d}_i \cdot \mathbf{d}_j^* = \delta_{ij}$, can be constructed according to

$$\Sigma = \left\{\mathbf{r} = \sum_{i=1}^{3+d} m_i\mathbf{d}_i \Big| m_i \in \mathbb{Z}\right\},$$

$$\mathbf{d}_i = \left(\mathbf{a}_i, -\sum_{j=1}^{d} \alpha_{ij}(1/c)\mathbf{e}_j\right),\ i = 1, \ldots, 3$$

$$\text{and } \mathbf{d}_{3+j} = \left(\mathbf{0}, (1/c)\mathbf{e}_j^*\right),\ j = 1, \ldots, d.$$

Consequently, the aperiodic structure in physical space $\mathbf{V}^{\parallel}$ is equivalent to a 3D section of the $(3 + d)$D hypercrystal.

4.6.3.1.1. Indexing

The 3D reciprocal space M^* of a $(3 + d)$D IMS consists of two separable contributions,

$$M^* = \left\{\mathbf{H} = \sum_{i=1}^{3} h_i\mathbf{a}_i^* + \sum_{j=1}^{d} m_j\mathbf{q}_j\right\},$$

the set of main reflections ($m_j = 0$) and the set of satellite reflections ($m_j \neq 0$) (Fig. 4.6.3.1). In most cases, the modulation is only a weak perturbation of the crystal structure. The main reflections are related to the average structure, the satellites to the difference between the average and actual structure. Consequently, the satellite reflections are generally much weaker than the main reflections and can be easily identified. Once the set of main reflections has been separated, a conventional basis $\mathbf{a}_i^*, i = 1, \ldots, 3$, for Λ^* is chosen.

The only ambiguity is in the assignment of rationally independent satellite vectors $\mathbf{q}_i$. They should be chosen inside the reciprocal-space unit cell (Brillouin zone) of Λ^* in such a way as to give a minimal number d of additional dimensions. If satellite vectors reach the Brillouin-zone boundary, centred $(3 + d)$D Bravais lattices are obtained. The star of satellite vectors has to be invariant under the point-symmetry group of the diffraction pattern. There should be no contradiction to a reasonable physical modulation model concerning period or propagation direction of the modulation wave. More detailed information on how to find the optimum basis and the correct setting is given by Janssen *et al.* (2004) and Janner *et al.* (1983*a*,*b*).

4.6.3.1.2. Diffraction symmetry

The Laue symmetry group $K^{\mathrm{L}} = \{R\}$ of the Fourier module M^*,

$$M^* = \left\{\mathbf{H} = \sum_{i=1}^{3} h_i\mathbf{a}_i^* + \sum_{j=1}^{d} m_j\mathbf{q}_j = \sum_{i=1}^{3+d} h_i\mathbf{a}_i^*\right\},\ \Lambda^* = \left\{\mathbf{H} = \sum_{i=1}^{3} h_i\mathbf{a}_i^*\right\},$$

is isomorphous to or a subgroup of one of the 11 3D crystallographic Laue groups leaving Λ^* invariant. The action of the point-group symmetry operators R on the reciprocal basis $\mathbf{a}_i^*, i = 1, \ldots, 3 + d$, can be written as

$$R\mathbf{a}_i^* = \sum_{j=1}^{3+d} \Gamma_{ij}^T(R)\mathbf{a}_j^*,\ i = 1, \ldots, 3 + d.$$

The $(3+d)\times(3+d)$ matrices $\Gamma^T(R)$ form a finite group of integral matrices which are reducible, since R is already an orthogonal transformation in 3D physical space. Consequently, R can be expressed as pair of orthogonal transformations $(R^{\parallel}, R^{\perp})$ in 3D physical and dD perpendicular space, respectively. Owing to their mutual orthogonality, no symmetry relationship exists between the set of main reflections and the set of satellite reflections. $\Gamma^T(R)$ is the transpose of $\Gamma(R)$ which acts on vector components in direct space.

For the $(3+d)$D direct-space (*superspace*) symmetry operator $(R_s, \mathbf{t}_s)$ and its matrix representation $\Gamma(R_s, \mathbf{t}_s)$ on Σ, the following decomposition can be performed:

$$\Gamma(R_s) = \begin{pmatrix} \Gamma^{\parallel}(R) & 0 \\ \Gamma^{M}(R) & \Gamma^{\perp}(R) \end{pmatrix} \text{ and } \mathbf{t}_s = (\mathbf{t}_3, \mathbf{t}_d).$$

$\Gamma^{\parallel}(R)$ is a 3×3 matrix, $\Gamma^{\perp}(R)$ is a $d\times d$ matrix and $\Gamma^{M}(R)$ is a $d\times 3$ matrix. The translation operator $\mathbf{t}_s$ consists of a 3D vector $\mathbf{t}_3$ and a dD vector $\mathbf{t}_d$. According to Janner & Janssen (1979), $\Gamma^{M}(R)$ can be derived from $\Gamma^{M}(R) = \sigma\Gamma^{\parallel}(R) - \Gamma^{\perp}(R)\sigma$. $\Gamma^{M}(R)$ has integer elements only as it contains components of primitive-lattice vectors of Λ^*, whereas σ in general consists of a rational and an irrational part: $\sigma = \sigma^{i} + \sigma^{r}$. Thus, only the rational part gives rise to nonzero entries in $\Gamma^{M}(R)$. With the order of the Laue group denoted by N, one obtains $\sigma^{i} \equiv (1/N)\sum_R \Gamma^{\perp}(R)\sigma\Gamma^{\parallel}(R)^{-1}$, where $\Gamma^{\perp}(R)\sigma^{i}\Gamma^{\parallel}(R)^{-1} = \sigma^{i}$, implying that $\Gamma^{M}(R) = \sigma^{r}\Gamma^{\parallel}(R) - \Gamma^{\perp}(R)\sigma^{r}$ and $0 = \sigma^{i}\Gamma^{\parallel}(R) - \Gamma^{\perp}(R)\sigma^{i}$.

Example

In the case of a 3D IMS with 1D modulation ($d = 1$) the $3\times d$ matrix

$$\sigma = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

has the components of the wavevector $\mathbf{q} = \sum_{i=1}^{3}\alpha_i \mathbf{a}_i^* = \mathbf{q}^{i} + \mathbf{q}^{r}$. $\Gamma^{\perp}(R) = \varepsilon = \pm 1$ because for $d = 1$, $\mathbf{q}$ can only be transformed into $\pm\mathbf{q}$. Corresponding to $\mathbf{q}^{i} \equiv (1/N)\sum_R \varepsilon R\mathbf{q}$, one obtains $R^T\mathbf{q}^{i} \equiv \varepsilon\mathbf{q}^{i}$ (modulo Λ^*). The 3×1 row matrix $\Gamma^{M}(R)$ is equivalent to the difference vector between $R^T\mathbf{q}$ and $\varepsilon\mathbf{q}$ (Janssen *et al.*, 2004).

For a monoclinic modulated structure with point group $2/m$ for M^* (unique axis $\mathbf{a}_3$) and satellite vector $\mathbf{q} = (1/2)\mathbf{a}_1^* + \alpha_3\mathbf{a}_3^*$, with α_3 an irrational number, one obtains

$$\begin{aligned}\mathbf{q}^{i} &\equiv (1/N)\sum_R \varepsilon R\mathbf{q} \\ &= \tfrac{1}{4}\left(+1\cdot\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1/2\\0\\\alpha_3\end{pmatrix} \right. \\ &\quad +1\cdot\begin{pmatrix}\bar1&0&0\\0&\bar1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1/2\\0\\\alpha_3\end{pmatrix} - 1\cdot\begin{pmatrix}\bar1&0&0\\0&\bar1&0\\0&0&\bar1\end{pmatrix}\begin{pmatrix}1/2\\0\\\alpha_3\end{pmatrix} \\ &\quad \left. -1\cdot\begin{pmatrix}1&0&0\\0&1&0\\0&0&\bar1\end{pmatrix}\begin{pmatrix}1/2\\0\\\alpha_3\end{pmatrix}\right) \\ &= \begin{pmatrix}0\\0\\\alpha_3\end{pmatrix}.\end{aligned}$$

From the relations $R^T\mathbf{q}^{i} \equiv \varepsilon\mathbf{q}^{i}$ (modulo Λ^*), it can be shown that the symmetry operations 1 and 2 are associated with the perpendicular-space transformations $\varepsilon = 1$, and m and $\bar1$ with $\varepsilon = -1$. The matrix $\Gamma^{M}(R)$ is given by

$$\begin{aligned}\Gamma^{M}(2) &= \sigma^{r}\Gamma^{\parallel}(2) - \Gamma^{\perp}(2)\sigma^{r} \\ &= \begin{pmatrix}1/2\\0\\0\end{pmatrix}\begin{pmatrix}\bar1&0&0\\0&\bar1&0\\0&0&1\end{pmatrix} - (+1)\begin{pmatrix}1/2\\0\\0\end{pmatrix} = \begin{pmatrix}\bar1\\0\\0\end{pmatrix}\end{aligned}$$

for the operation 2, for instance.

The matrix representations $\Gamma^T(R_s)$ of the symmetry operators R in reciprocal $(3+d)$D superspace decompose according to

$$\Gamma^T(R_s) = \begin{pmatrix} \Gamma^{\parallel T}(R) & \Gamma^{MT}(R) \\ 0 & \Gamma^{\perp T}(R) \end{pmatrix}.$$

Phase relationships between modulation functions of symmetry-equivalent atoms can give rise to systematic extinctions of different classes of satellite reflections. The extinction rules may include indices of both main and satellite reflections. A full list of systematic absences is given in the table of $(3+1)$D superspace groups (Janssen *et al.*, 2004). Thus, once point symmetry and systematic absences are found, the superspace group can be obtained from the tables in a way analogous to that used for regular 3D crystals. A different approach for the symmetry description of IMSs from the 3D Fourier-space perspective has been given by Dräger & Mermin (1996).

4.6.3.1.3. Structure factor

The structure factor of a periodic structure is defined as the Fourier transform of the density distribution $\rho(\mathbf{r})$ of its unit cell (UC):

$$F(\mathbf{H}) = \int_{\mathrm{UC}} \rho(\mathbf{r})\exp(2\pi i\mathbf{H}\cdot\mathbf{r})\ \mathrm{d}\mathbf{r}.$$

The same is valid in the case of the $(3+d)$D description of IMSs. The parallel- and perpendicular-space components are orthogonal to each other and can be separated. The Fourier transform of the parallel-space component of the electron-density distribution of a single atom gives the usual atomic scattering factors $f_k(\mathbf{H}^{\parallel})$. For the structure-factor calculation, one does not need to use $\rho(\mathbf{r})$ explicitly. The hyperatoms correspond to the convolution of the electron-density distribution in 3D physical space with the modulation function in dD perpendicular space. Therefore, the Fourier transform of the $(3+d)$D hyperatoms is simply the product of the Fourier transform $f_k(\mathbf{H}^{\parallel})$ of the physical-space component with the Fourier transform of the perpendicular-space component, the modulation function.

For a general *displacive modulation* one obtains for the ith coordinate x_{ik} of the kth atom in 3D physical space

$$x_{ik} = \bar{x}_{ik} + u_{ik}(\bar{x}_4, \ldots, \bar{x}_{3+d}),\ \ i = 1, \ldots, 3,$$

where $\bar{x}_{ik}$ are the basic-structure coordinates and $u_{ik}(\bar{x}_4, \ldots, \bar{x}_{3+d})$ are the modulation functions with unit periods in their arguments (Fig. 4.6.3.2). The arguments are $\bar{x}_{3+j} = \alpha_{ij}\bar{x}_{ik}^0 + t_j,\ j = 1, \ldots, d$, where $\bar{x}_{ik}^0$ are the coordinates of the kth atom referred to the origin of its unit cell and t_j are the phases of the modulation functions. The modulation functions $u_{ik}(\bar{x}_4, \ldots, \bar{x}_{3+d})$ themselves can be expressed in terms of a Fourier series as

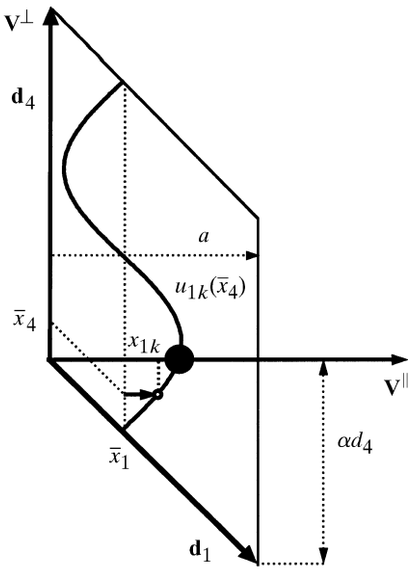


Fig. 4.6.3.2. The relationships between the coordinates $x_{1k}, x_{4k}, \bar{x}_1, \bar{x}_4$ and the modulation function u_{1k} in a special section of the $(3+d)$D space.

$$u_{ik}(\bar{x}_4, \ldots, \bar{x}_{3+d}) = \sum_{n_1=1}^{\infty} \ldots \sum_{n_d=1}^{\infty} \{{}^{u}C_{ik}^{n_1 \ldots n_d} \cos[2\pi(n_1\bar{x}_4 + \ldots + n_d\bar{x}_{3+d})] + {}^{u}S_{ik}^{n_1 \ldots n_d} \sin[2\pi(n_1\bar{x}_4 + \ldots + n_d\bar{x}_{3+d})]\},$$

where n_j are the orders of harmonics for the jth modulation wave of the ith component of the kth atom and their amplitudes are ${}^{u}C_{ik}^{n_1 \ldots n_d}$ and ${}^{u}S_{ik}^{n_1 \ldots n_d}$.

Analogous expressions can be derived for a *density modulation*, *i.e.*, the modulation of the occupation probability $p_k(\bar{x}_4, \ldots, \bar{x}_{3+d})$:

$$p_k(\bar{x}_4, \ldots, \bar{x}_{3+d}) = \sum_{n_1=1}^{\infty} \ldots \sum_{n_d=1}^{\infty} \{{}^{p}C_{k}^{n_1 \ldots n_d} \cos[2\pi(n_1\bar{x}_4 + \ldots + n_d\bar{x}_{3+d})] + {}^{p}S_{k}^{n_1 \ldots n_d} \sin[2\pi(n_1\bar{x}_4 + \ldots + n_d\bar{x}_{3+d})]\},$$

and for the modulation of the tensor of thermal parameters $B_{ijk}(\bar{x}_4, \ldots, \bar{x}_{3+d})$:

$$B_{ijk}(\bar{x}_4, \ldots, \bar{x}_{3+d}) = \sum_{n_1=1}^{\infty} \ldots \sum_{n_d=1}^{\infty} \{{}^{B}C_{ijk}^{n_1 \ldots n_d} \cos[2\pi(n_1\bar{x}_4 + \ldots + n_d\bar{x}_{3+d})] + {}^{B}S_{ijk}^{n_1 \ldots n_d} \sin[2\pi(n_1\bar{x}_4 + \ldots + n_d\bar{x}_{3+d})]\}.$$

The resulting structure-factor formula is

$$F(\mathbf{H}) = \sum_{k=1}^{N'} \sum_{(R,t)} \int_0^1 \mathrm{d}\bar{x}_{4,k} \ldots \int_0^1 \mathrm{d}\bar{x}_{3+d,k} f_k(\mathbf{H}^{\parallel}) p_k \times \exp\left(-\sum_{i,j=1}^{3+d} h_i[RB_{ijk}R^T]h_j + 2\pi i \sum_{j=1}^{3+d} h_j R x_{jk} + h_j t_j\right)$$

for summing over the set $(R, \mathbf{t})$ of superspace symmetry operations and the set of N' atoms in the asymmetric unit of the $(3+d)$D unit cell (Yamamoto, 1982). Different approaches without numerical integration based on analytical expressions including Bessel functions have also been developed. For more information see Paciorek & Chapuis (1994), Petricek, Maly & Cisarova (1991), and references therein.

For illustration, some fundamental IMSs will be discussed briefly (see Korekawa, 1967; Böhm, 1977).

Harmonic density modulation. A harmonic density modulation can result on average from an ordered distribution of vacancies on atomic positions. For an IMS with N atoms per unit cell one obtains for a harmonic modulation of the occupancy factor

$$p_k = (p_k^0/2)\{1 + \cos[2\pi(\bar{x}_{4,k} + \varphi_k)]\}, \quad 0 \le p_k^0 \le 1,$$

the structure-factor formula for the mth order satellite $(0 \le m \le 1)$

$$F_0(\mathbf{H}) = (1/2)\sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_k),$$
$$F_m(\mathbf{H}) = (1/2)\sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}) (p_k^0/2)^{|m|} \exp\left[2\pi i\left(\sum_{i=1}^{3} h_i x_{ik} + m\varphi_k\right)\right].$$

Thus, a linear correspondence exists between the structure-factor magnitudes of the satellite reflections and the amplitude of the density modulation. Furthermore, only first-order satellites exist, since the modulation wave consists only of one term. An important criterion for the existence of a density modulation is that a pair of satellites around the origin of the reciprocal lattice exists (Fig. 4.6.3.3).

Symmetric rectangular density modulation. The box-function-like modulated occupancy factor can be expanded into a Fourier series,

$$p_k = p_k^0(4/\pi)\left\{\sum_{n=1}^{\infty}[(-1)^{n+1}/(2n-1)]\cos[2\pi(2n-1)(\bar{x}_{4,k} + \varphi_k)]\right\}, \quad 0 \le p_k^0 \le 1,$$

and the resulting structure factor of the mth order satellite is

$$F_0(\mathbf{H}) = (1/2)\sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}) \exp\left(2\pi i \sum_{i=1}^{3} h_i x_{ik}\right),$$
$$F_m(\mathbf{H}) = (1/\pi m)\sin(m\pi/2)\sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}) p_k^0 \times \exp\left[2\pi i\left(\sum_{i=1}^{3} h_i x_{ik} + m\varphi_k\right)\right].$$

According to this formula, only odd-order satellites occur in the diffraction pattern. Their structure-factor magnitudes decrease linearly with the order $|m|$ (Fig. 4.6.3.3*b*).

Harmonic displacive modulation. The displacement of the atomic coordinates is given by the function

$$x_{ik} = x_{ik}^0 + A_{ik}\cos[2\pi(\bar{x}_{4,k} + \varphi_k)], \quad i = 1, \ldots, 3,$$

and the structure factor by

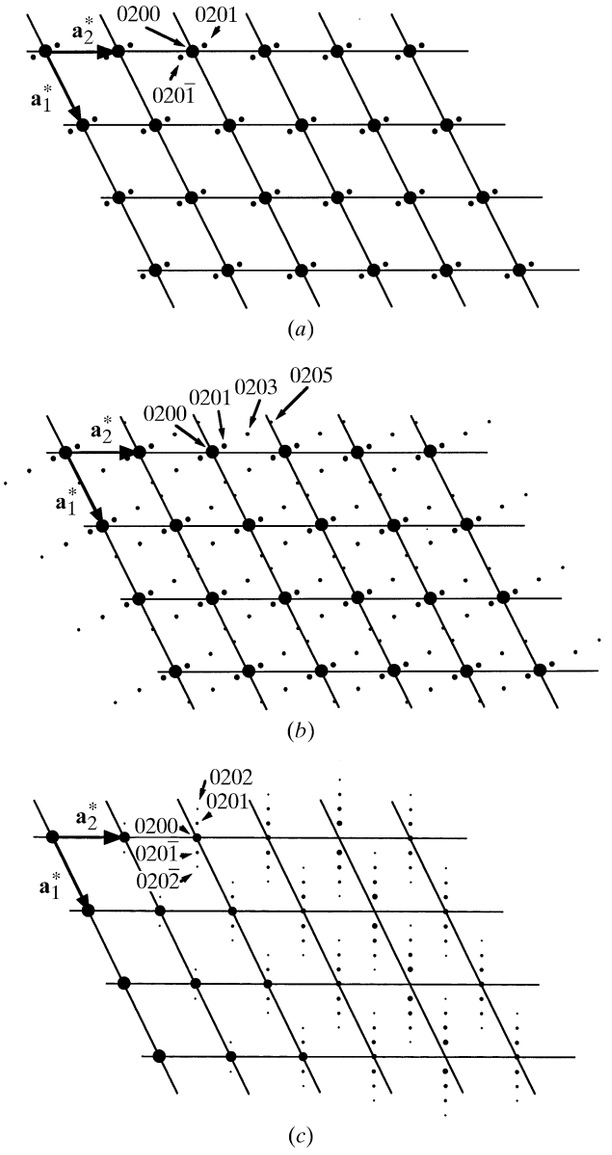


Fig. 4.6.3.3. Schematic diffraction patterns for 3D IMSs with (*a*) 1D harmonic and (*b*) rectangular density modulation. The modulation direction is parallel to $\mathbf{a}_2$. In (*a*) only first-order satellites exist; in (*b*), all odd-order satellites can be present. In (*c*), the diffraction pattern of a harmonic displacive modulation along $\mathbf{a}_1$ with amplitudes parallel to $\mathbf{a}_2^*$ is depicted. Several reflections are indexed. The areas of the circles are proportional to the reflection intensities.

$$F_0(\mathbf{H}) = \sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}) J_0(2\pi \mathbf{H}^{\parallel} \cdot \mathbf{A}_k) \exp\left(2\pi i \sum_{i=1}^{3} h_i x_{ik}\right),$$

$$F_m(\mathbf{H}) = \sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}) J_m(2\pi \mathbf{H}^{\parallel} \cdot \mathbf{A}_k) \times \exp\left[2\pi i \left(\sum_{i=1}^{3} h_i x_{ik} + m\varphi_k\right)\right].$$

The structure-factor magnitudes of the *m*th-order satellite reflections are a function of the *m*th-order Bessel functions. The arguments of the Bessel functions are proportional to the scalar products of the amplitude and the diffraction vector. Consequently, the intensity of the satellites will vary characteristically as a function of the length of the diffraction vector. Each main reflection is accompanied by an infinite number of satellite reflections (Figs. 4.6.3.3*c* and 4.6.3.4).

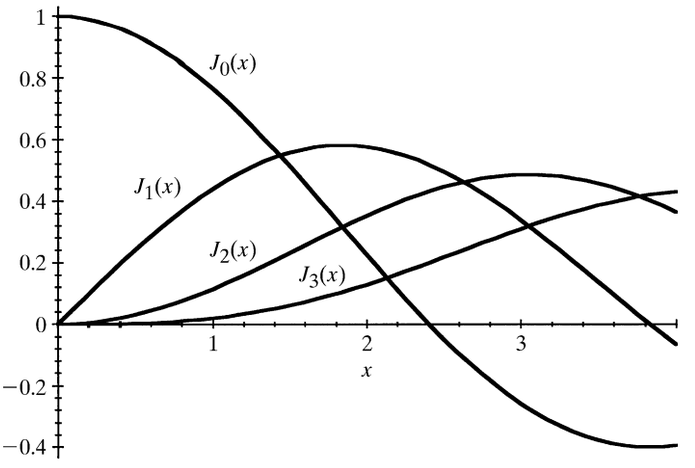


Fig. 4.6.3.4. The relative structure-factor magnitudes of *m*th-order satellite reflections for a harmonic displacive modulation are proportional to the values of the *m*th-order Bessel function $J_m(x)$.

4.6.3.2. Composite structures (CSs)

Composite structures consist of N mutually incommensurate substructures with N basic sublattices $\Lambda_\nu = \{\mathbf{a}_{1\nu}, \mathbf{a}_{2\nu}, \mathbf{a}_{3\nu}\}$, with $\nu = 1, \ldots, N$. The reciprocal sublattices $\Lambda_\nu^* = \{\mathbf{a}_{1\nu}^*, \mathbf{a}_{2\nu}^*, \mathbf{a}_{3\nu}^*\}$, with $\nu = 1, \ldots, N$, have either only the origin of the reciprocal lattice or one or two reciprocal-lattice directions in common. Thus, one needs $(3 + d) < 3N$ reciprocal-basis vectors for integer indexing of diffraction patterns that show Bragg reflections at positions given by the Fourier module M^*. The CSs discovered to date have at least one lattice direction in common and consist of a maximum number of $N = 3$ substructures. They can be divided in three main classes: channel structures, columnar packings and layer packings (see van Smaalen, 1992, 1995).

In the following, the approach of Janner & Janssen (1980*b*) and van Smaalen (1992, 1995, and references therein) for the description of CSs is used. The set of diffraction vectors of a CS, *i.e.* its Fourier module $M^* = \{\sum_{i=1}^{3+d} h_i \mathbf{a}_i^*\}$, can be split into the contributions of the ν subsystems by employing $3 \times (3 + d)$ matrices $Z_{ik\nu}$ with integer coefficients $\mathbf{a}_{i\nu}^* = \sum_{k=1}^{3+d} Z_{ik\nu} \mathbf{a}_k^*$, $i = 1, \ldots, 3$. In the general case, each subsystem will be modulated with the periods of the others due to their mutual interactions. Thus, in general, CSs consist of several intergrown incommensurately modulated substructures. The satellite vectors $\mathbf{q}_{j\nu}$, $j = 1, \ldots, d$, referred to the νth subsystem can be obtained from M^* by applying the $d \times (3 + d)$ integer matrices $V_{jk\nu}$: $\mathbf{q}_{j\nu} = \sum_{k=1}^{3+d} V_{jk\nu} \mathbf{a}_k^*$, $j = 1, \ldots, d$. The matrices consisting of the components σ_ν of the satellite vectors $\mathbf{q}_{j\nu}$ with regard to the reciprocal sublattices Λ_ν^* can be calculated by $\sigma_\nu = (V_{3\nu} + V_{d\nu}\sigma)(Z_{3\nu} + Z_{d\nu}\sigma)^{-1}$, where the subscript 3 refers to the 3×3 submatrix of physical space and the subscript d to the $d \times d$ matrix of the internal space. The juxtaposition of the $3 \times (3 + d)$ matrix Z_ν and the $d \times (3 + d)$ matrix V_ν defines the non-singular $(3 + d) \times (3 + d)$ matrix W_ν,

$$W_\nu = \begin{pmatrix} Z_\nu \\ V_\nu \end{pmatrix}.$$

This matrix allows the reinterpretation of the Fourier module M^* as the Fourier module $M_\nu^* = M^* W_\nu$ of a d-dimensionally modulated subsystem ν. It also describes the coordinate transformation between the superspace basis Σ and Σ_ν.

The superspace description is obtained analogously to that for IMSs (see Section 4.6.3.1) by considering the 3D Fourier module M^* of rank $3 + d$ as the projection of a $(3 + d)$D reciprocal lattice Σ^* upon the physical space. Thus, one obtains for the definition

of the direct and reciprocal $(3+d)$ lattices (Janner & Janssen, 1980*b*)

$$\Sigma^*: \begin{cases} \mathbf{a}_i^* & = (\mathbf{a}_i^*, \mathbf{0}) & i = 1, \ldots, 3 \\ \mathbf{a}_{3+j}^* & = (\mathbf{a}_{3+j}^*, \mathbf{e}_j^*) & j = 1, \ldots, d \end{cases}$$

$$\Sigma: \begin{cases} \mathbf{a}_i & = (\mathbf{a}_i, -\sum_{j=1}^{d} \sigma_{ji} \mathbf{e}_j) & i = 1, \ldots, 3 \\ \mathbf{a}_{3+j} & = (\mathbf{0}, \mathbf{e}_j) & j = 1, \ldots, d. \end{cases}$$

4.6.3.2.1. Indexing

The indexing of diffraction patterns of composite structures can be performed in the following way:

(1) find the minimum number of reciprocal lattices Λ_ν^* necessary to index the diffraction pattern;

(2) find a basis for M^*, the union of sublattices Λ_ν^*;

(3) find the appropriate superspace embedding.

The $(3+d)$ vectors $\mathbf{a}_i^*$ forming a basis for the 3D Fourier module $M^* = \{\sum_{i=1}^{3+d} h_i \mathbf{a}_i^*\}$ can be chosen such that $\mathbf{a}_1^*$, $\mathbf{a}_2^*$ and $\mathbf{a}_3^*$ are linearly independent. Then the remaining d vectors can be described as a linear combination of the first three, defining the $d \times 3$ matrix σ: $\mathbf{a}_{3+j}^* = \sum_{i=1}^{3+d} \sigma_{ji} \mathbf{a}_i^*$, $j = 1, \ldots, d$. This is formally equivalent to the reciprocal basis obtained for an IMS (see Section 4.6.3.1) and one can proceed in an analogous way to that for IMSs.

4.6.3.2.2. Diffraction symmetry

The symmetry of CSs can be described with basically the same formalism as used for IMSs. This is a consequence of the formally equivalent applicability of the higher-dimensional approach, in particular of the superspace-group theory developed for IMSs [see Janner & Janssen (1980*a*,*b*); van Smaalen (1991, 1992); Yamamoto (1992*a*)].

4.6.3.2.3. Structure factor

The structure factor $F(\mathbf{H})$ of a composite structure consists of the weighted contributions of the subsystem structure factors $F_\nu(\mathbf{H}_\nu)$:

$$F(\mathbf{H}) = \sum_\nu |J_\nu| F_\nu(\mathbf{H}_\nu);$$

$$F_\nu(\mathbf{H}) = \sum_{(R^\nu, \mathbf{t}^\nu)}^{\infty} \sum_{k=1}^{N'} \int_0^1 \mathrm{d}\bar{x}_{4,k}^\nu \ldots \int_0^1 \mathrm{d}\bar{x}_{3+d,k}^\nu f_k^\nu(\mathbf{H}^\parallel) p_k^\nu \times \exp\left(- \sum_{i,j=1}^{3+d} h_i^\nu [R^\nu B_{ijk}^\nu R^{\nu T}] h_j^\nu + 2\pi i \sum_{j=1}^{3+d} h_j^\nu R^\nu x_{jk}^\nu + h_j^\nu t_j^\nu \right),$$

with coefficients similar to those for IMSs.

The weights are the Jacobians of the transformations from $\mathbf{t}_\nu$ to $\mathbf{t}$, and $\mathbf{H}_\nu$ are the reflection indices with respect to the subsystem Fourier modules M_ν^* (van Smaalen, 1995, and references therein). The relative values of $|J_\nu|$, where $J_\nu = \det\left[(V_{d\nu} - \sigma_\nu \cdot Z_{d\nu})^{-1}\right]$, are related to the volume ratios of the contributing subsystems. The subsystem structure factors correspond to those for IMSs (see Section 4.6.3.1). Besides this formula, based on the publications of Yamamoto (1982) and van Smaalen (1995), different structure-factor equations have been discussed (Kato, 1990; Petricek, Maly, Coppens *et al.*, 1991).

4.6.3.3. Quasiperiodic structures (QSs)

4.6.3.3.1. 3D structures with 1D quasiperiodic order

Structures with quasiperiodic order in one dimension and lattice symmetry in the other two dimensions are the simplest representatives of quasicrystals. A few phases of this structure type have been identified experimentally (see Steurer, 1990). Since the Fibonacci chain represents the most important model of a 1D quasiperiodic structure, it will be used in this section to represent the quasiperiodic direction of 3D structures with 1D quasiperiodic order. As discussed in Section 4.6.2.4, 1D quasiperiodic structures are on the borderline between quasiperiodic and incommensurately modulated structures. They can be described using either of the two approaches. In the following, the quasiperiodic description will be preferred to take account of the scaling symmetry.

The electron-density-distribution function $\rho(\mathbf{r})$ of a 1D quasiperiodically ordered 3D crystal can be represented by a Fourier series:

$$\rho(\mathbf{r}) = (1/V) \sum_{\mathbf{H}} F(\mathbf{H}) \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}).$$

The Fourier coefficients (*structure factors*) $F(\mathbf{H})$ differ from zero only for reciprocal-space vectors $\mathbf{H} = \sum_{i=1}^{3} h_i^\parallel \mathbf{a}_i^*$ with ${h_1}^\parallel \in \mathbb{R}$, ${h_2}^\parallel, {h_3}^\parallel \in \mathbb{Z}$ or with integer indexing $\mathbf{H} = \sum_{i=1}^{4} h_i \mathbf{a}_i^*$ with $h_i \in \mathbb{Z}$. The set of all vectors $\mathbf{H}$ forms a Fourier module $M^* = \{\mathbf{H}^\parallel = \sum_{i=1}^{4} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ of rank 4 which can be decomposed into two rank 2 submodules $M^* = M_1^* \oplus M_2^*$. $M_1^* = \{h_1 \mathbf{a}_1^* + h_2 \mathbf{a}_2^*\}$ corresponds to a $\mathbb{Z}$ module of rank 2 in a 1D subspace, $M_2^* = \{h_3 \mathbf{a}_3^* + h_4 \mathbf{a}_4^*\}$ corresponds to a $\mathbb{Z}$ module of rank 2 in a 2D subspace. Consequently, the first submodule can be considered as a projection from a 2D reciprocal lattice, $M_1^* = \pi^\parallel(\Sigma^*)$, while the second submodule is of the form of a reciprocal lattice, $M_2^* = \Lambda^*$.

Hence, the reciprocal-basis vectors $\mathbf{a}_i^*$, $i = 1, \ldots, 4$, can be considered to be projections of reciprocal-basis vectors $\mathbf{d}_i^*$, $i = 1, \ldots, 4$, spanning a 4D reciprocal lattice, onto the physical space $\Sigma^* = \{\mathbf{H} = \sum_{i=1}^{4} h_i \mathbf{d}_i^* | h_i \in \mathbb{Z}\}$, with

$$\mathbf{d}_1^* = a_1^* \begin{pmatrix} 1 \\ -\tau \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{d}_2^* = a_1^* \begin{pmatrix} \tau \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{d}_3^* = a_3^* \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{d}_4^* = a_4^* \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

A direct lattice Σ with basis $\mathbf{d}_i$, $i = 1, \ldots, 4$ and $\mathbf{d}_i \cdot \mathbf{d}_j^* = \delta_{ij}$, can be constructed according to (compare Fig. 4.6.2.8) $\Sigma = \{\mathbf{r} = \sum_{i=1}^{4} m_i \mathbf{d}_i | m_i \in \mathbb{Z}\}$, with

$$\mathbf{d}_1 = \frac{1}{a_1^*(2+\tau)} \begin{pmatrix} 1 \\ -\tau \\ 0 \\ 0 \end{pmatrix}, \mathbf{d}_2 = \frac{1}{a_1^*(2+\tau)} \begin{pmatrix} \tau \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$\mathbf{d}_3^* = \frac{1}{a_3^*} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{d}_4^* = \frac{1}{a_4^*} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Consequently, the structure in physical space $\mathbf{V}^\parallel$ is equivalent to a 3D section of the 4D hypercrystal.

4.6.3.3.1.1. Indexing

The reciprocal space of the Fibonacci chain is densely filled with Bragg reflections (Figs. 4.6.2.9 and 4.6.3.5). According to the nD embedding method, the shorter the parallel-space distance $\Delta\mathbf{H}^\parallel = \mathbf{H}_2^\parallel - \mathbf{H}_1^\parallel$ between two Bragg reflections, the larger the corresponding perpendicular-space distance $\Delta\mathbf{H}^\perp = \mathbf{H}_2^\perp - \mathbf{H}_1^\perp$ becomes. Since the structure factor $F(\mathbf{H})$ decreases rapidly as a

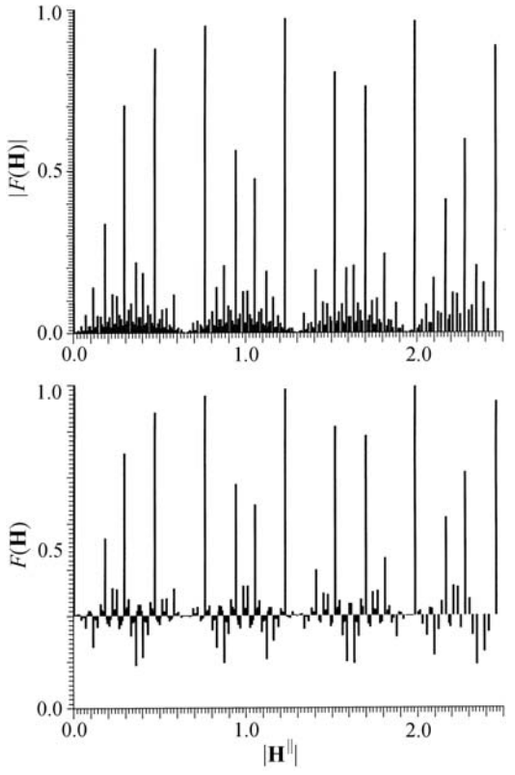


Fig. 4.6.3.5. The structure factors $F(\mathbf{H})$ (below) and their magnitudes $|F(\mathbf{H})|$ (above) of a Fibonacci chain decorated with equal point atoms are shown as a function of the parallel-space component $|\mathbf{H}^{\parallel}|$ of the diffraction vector. The short distance in the Fibonacci chain is S = 2.5 Å, all structure factors within $0 \leq |\mathbf{H}| \leq 2.5$ Å^{-1} have been calculated and normalized to $F(00) = 1$.

function of $\mathbf{H}^{\perp}$ (Fig. 4.6.3.6), 'neighbouring' reflections of strong Bragg peaks are extremely weak and, consequently, the reciprocal space appears to be filled with *discrete* Bragg peaks even for low-resolution experiments.

This property allows an unambiguous identification of a correct set of reciprocal-basis vectors. However, infinitely many sets allowing a correct indexing of the diffraction pattern with integer indices exist. Nevertheless, an optimum basis (low indices are assigned to strong reflections) can be derived: the intensity distribution, not the metrics, characterizes the best choice of indexing. Once the minimum distance S in the structure is identified from chemical considerations, the reciprocal basis should be chosen as described in Section 4.6.2.4. It has to be kept in mind, however, that the identification of the metrics is not sufficient to distinguish in the 1D aperiodic case between an incommensurately modulated structure, a quasiperiodic structure or special kinds of structures with fractally shaped atomic surfaces.

A correct set of reciprocal-basis vectors can be identified in the following way:

(1) Find pairs of strong reflections whose physical-space diffraction vectors are related to each other by the factor τ.

(2) Index these reflections by assigning an appropriate value to a^*. This value should be derived from the shortest interatomic distance S expected in the structure.

(3) The reciprocal basis is correct if all observable Bragg reflections can be indexed with integer numbers.

4.6.3.3.1.2. Diffraction symmetry

The possible Laue symmetry group K^{3D} of the Fourier module $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{4} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ is any one of the direct product $K^{3D} = K^{2D} \otimes K^{1D} \otimes \bar{1}$. K^{2D} corresponds to one of the ten crystallographic 2D point groups, $K^{1D} = \{1\}$ in the general case of a quasiperiodic stacking of periodic layers. Consequently, the nine Laue groups $\bar{1}$, $2/m$, mmm, $4/m$, $4/mmm$, $\bar{3}$, $\bar{3}m$, $6/m$ and $6/mmm$ are possible. These are all 3D crystallographic Laue groups except for the two cubic ones.

The (unweighted) Fourier module shows only 2D lattice symmetry. In the third dimension, the submodule M_1^* remains invariant under the scaling symmetry operation $S^n M_1^* = \tau^n M_1^*$ with $n \in \mathbb{Z}$. The scaling symmetry operators S^n form an infinite group $s = \{\ldots, S^{-1}, S^0, S^1, \ldots\}$ of reciprocal-basis transformations S^n in superspace,

$$S^n = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_D^n, \quad S^{-1} = \begin{pmatrix} \bar{1} & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_D,$$

$$S^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_D,$$

and act on the reciprocal basis $\mathbf{d}_i^*$ in superspace.

4.6.3.3.1.3. Structure factor

The structure factor of a periodic structure is defined as the Fourier transform of the density distribution $\rho(\mathbf{r})$ of its unit cell (UC):

$$F(\mathbf{H}) = \int_{\mathrm{UC}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}) \, \mathrm{d}\mathbf{r}.$$

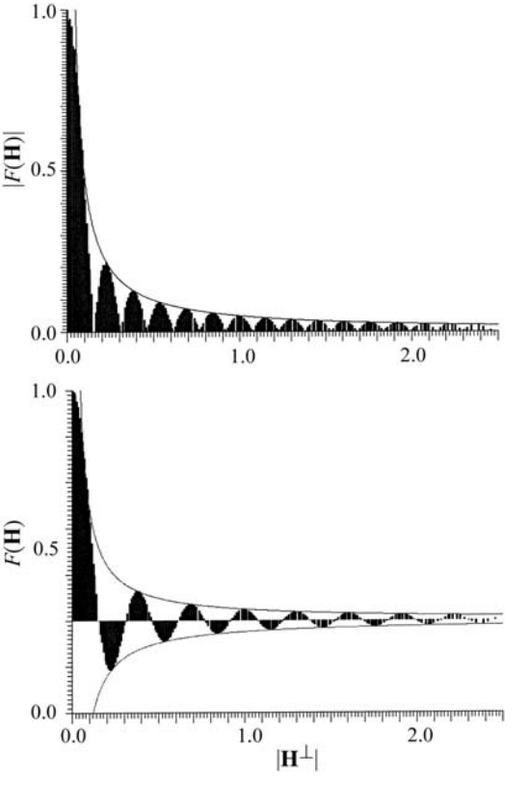


Fig. 4.6.3.6. The structure factors $F(\mathbf{H})$ (below) and their magnitudes $|F(\mathbf{H})|$ (above) of a Fibonacci chain decorated with equal point atoms are shown as a function of the perpendicular-space component $|\mathbf{H}^{\perp}|$ of the diffraction vector. The short distance in the Fibonacci chain is S = 2.5 Å, all structure factors within $0 \leq |\mathbf{H}| \leq 2.5$ Å^{-1} have been calculated and normalized to $F(00) = 1$.

The same is valid in the case of the nD description of a quasiperiodic structure. The parallel- and perpendicular-space components are orthogonal to each other and can be separated. In the case of the 1D Fibonacci sequence, the Fourier transform of the parallel-space component of the electron-density distribution of a single atom gives the usual atomic scattering factor $f(\mathbf{H}^{\parallel})$. Parallel to $x^{\perp}$, $\rho(\mathbf{r})$ adopts values $\neq 0$ only within the interval $-(1+\tau)/[2a^*(2+\tau)] \leq x^{\perp} \leq (1+\tau)/[2a^*(2+\tau)]$ and one obtains

$$F(\mathbf{H}) = f(\mathbf{H}^{\parallel})[a^*(2+\tau)]/(1+\tau) \times \int_{-(1+\tau)/[2a^*(2+\tau)]}^{+(1+\tau)/[2a^*(2+\tau)]} \exp(2\pi i \mathbf{H}^{\perp} \cdot x^{\perp})\, \mathrm{d}x^{\perp}.$$

The factor $[a^*(2+\tau)]/(1+\tau)$ results from the normalization of the structure factors to $F(\mathbf{0}) = f(0)$. With

$$\mathbf{H} = h_1\mathbf{d}_1^* + h_2\mathbf{d}_2^* + h_3\mathbf{d}_3^* + h_4\mathbf{d}_4^* = h_1a_1^*\begin{pmatrix}1\\-\tau\\0\\0\end{pmatrix} + h_2a_1^*\begin{pmatrix}\tau\\1\\0\\0\end{pmatrix} + h_3a_3^*\begin{pmatrix}0\\0\\1\\0\end{pmatrix} + h_4a_4^*\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

and $\mathbf{H}^{\perp} = a_1^*(-\tau h_1 + h_2)$ the integrand can be rewritten as

$$F(\mathbf{H}) = f(\mathbf{H}^{\parallel})[a^*(2+\tau)]/(1+\tau) \times \int_{-(1+\tau)/[2a^*(2+\tau)]}^{+(1+\tau)/[2a^*(2+\tau)]} \exp[2\pi i(-\tau h_1 + h_2)x^{\perp}]\, \mathrm{d}x^{\perp},$$

yielding

$$F(\mathbf{H}) = f(\mathbf{H}^{\parallel})(2+\tau)/[2\pi i(-\tau h_1 + h_2)(1+\tau)] \times \exp[2\pi i(-\tau h_1 + h_2)x^{\perp}]\Big|_{-(1+\tau)/[2a^*(2+\tau)]}^{+(1+\tau)/[2a^*(2+\tau)]}.$$

Using $\sin x = (\mathrm{e}^{ix} - \mathrm{e}^{-ix})/2i$ gives

$$F(\mathbf{H}) = f(\mathbf{H}^{\parallel})(2+\tau)/[\pi(-\tau h_1 + h_2)(1+\tau)] \times \sin[\pi(1+\tau)(-\tau h_1 + h_2)]/(2+\tau).$$

Thus, the structure factor has the form of the function $\sin(x)/x$ with x a perpendicular reciprocal-space coordinate. The upper and lower limiting curves of this function are given by the hyperbolae $\pm 1/x$ (Fig. 4.6.3.6). The continuous shape of $F(\mathbf{H})$ as a function of $\mathbf{H}^{\perp}$ allows the estimation of an overall temperature factor and atomic scattering factor for reflection-data normalization (compare Figs. 4.6.3.6 and 4.6.3.7).

In the case of a 3D crystal structure which is quasiperiodic in one direction, the structure factor can be written in the form

$$F(\mathbf{H}) = \sum_{k=1}^{n}\big[T_k(\mathbf{H})f_k(\mathbf{H}^{\parallel})g_k(\mathbf{H}^{\perp})\exp(2\pi i\mathbf{H}\cdot\mathbf{r}_k)\big].$$

The sum runs over all n averaged hyperatoms in the 4D unit cell of the structure. The *geometric form factor* $g_k(\mathbf{H}^{\perp})$ corresponds to the Fourier transform of the kth atomic surface,

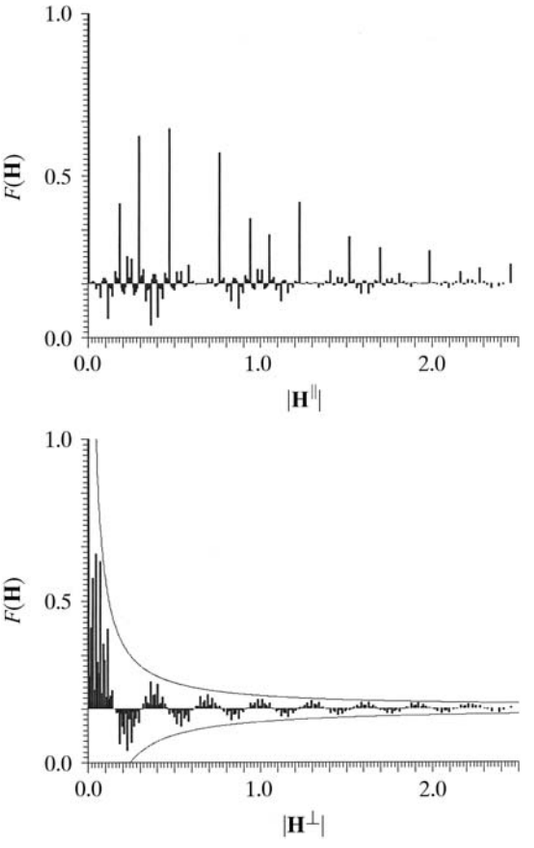


Fig. 4.6.3.7. The structure factors $F(\mathbf{H})$ of the Fibonacci chain decorated with aluminium atoms (U_{overall} = 0.005 Å^2) as a function of the parallel (above) and the perpendicular (below) component of the diffraction vector. The short distance is S = 2.5 Å, all structure factors within $0 \leq |\mathbf{H}| \leq 2.5$ Å^{-1} have been calculated and normalized to $F(00) = 1$.

$$g_k(\mathbf{H}^{\perp}) = (1/A_{\mathrm{UC}}^{\perp})\int_{A_k}\exp(2\pi i\mathbf{H}^{\perp}\cdot\mathbf{r}^{\perp})\ \mathrm{d}\mathbf{r}^{\perp},$$

normalized to $A_{\mathrm{UC}}^{\perp}$, the area of the 2D unit cell projected upon $\mathbf{V}^{\perp}$, and A_k, the area of the kth atomic surface.

The atomic temperature factor $T_k(\mathbf{H})$ can also have perpendicular-space components. Assuming only harmonic (static or dynamic) displacements in parallel and perpendicular space one obtains, in analogy to the usual expression (Willis & Pryor, 1975),

$$T_k(\mathbf{H}) = T_k(\mathbf{H}^{\parallel}, \mathbf{H}^{\perp}) = \exp(-2\pi^2\mathbf{H}^{\parallel T}\langle\mathbf{u}_i^{\parallel}\mathbf{u}_j^{\parallel T}\rangle\mathbf{H}^{\parallel})\exp(-2\pi^2\mathbf{H}^{\perp T}\langle\mathbf{u}_i^{\perp}\mathbf{u}_j^{\perp T}\rangle\mathbf{H}^{\perp}),$$

with

$$\langle\mathbf{u}_i^{\parallel}\mathbf{u}_j^{\parallel T}\rangle = \begin{pmatrix}\langle\mathbf{u}_1^{\parallel 2}\rangle & \langle\mathbf{u}_1^{\parallel}\cdot\mathbf{u}_2^{\parallel T}\rangle & \langle\mathbf{u}_1^{\parallel}\cdot\mathbf{u}_3^{\parallel T}\rangle\\ \langle\mathbf{u}_2^{\parallel}\cdot\mathbf{u}_1^{\parallel T}\rangle & \langle\mathbf{u}_2^{\parallel 2}\rangle & \langle\mathbf{u}_2^{\parallel}\cdot\mathbf{u}_3^{\parallel T}\rangle\\ \langle\mathbf{u}_3^{\parallel}\cdot\mathbf{u}_1^{\parallel T}\rangle & \langle\mathbf{u}_3^{\parallel}\cdot\mathbf{u}_2^{\parallel T}\rangle & \langle\mathbf{u}_3^{\parallel 2}\rangle\end{pmatrix}$$

and $\langle\mathbf{u}_i^{\perp}\mathbf{u}_j^{\perp T}\rangle = \langle\mathbf{u}_4^{\perp}\rangle$.

The elements of the type $\langle\mathbf{u}_i\cdot\mathbf{u}_j^T\rangle$ represent the average values of the atomic displacements along the ith axis times the displacement along the jth axis on the V basis.

4.6.3.3.1.4. Intensity statistics

In the following, only the properties of the quasiperiodic component of the 3D structure, namely the Fourier module M_1^*,

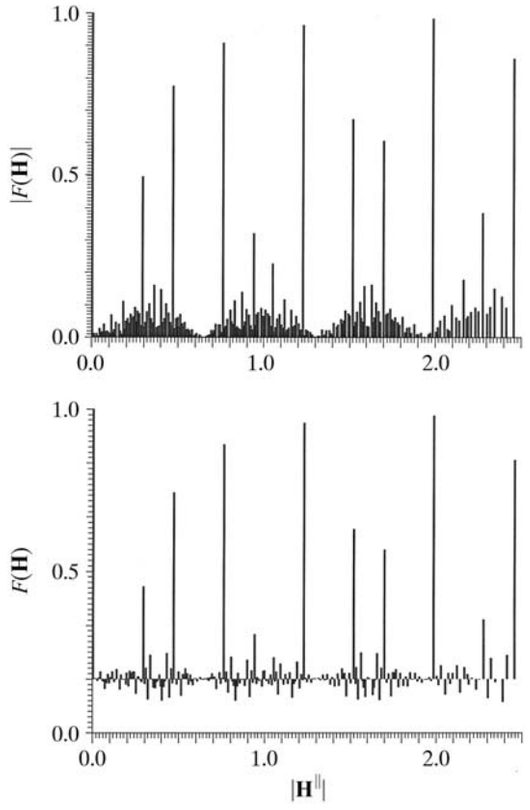


Fig. 4.6.3.8. The structure factors $F(\mathbf{H})$ (below) and their magnitudes $|F(\mathbf{H})|$ (above) of the squared Fibonacci chain decorated with equal point atoms are shown as a function of the parallel-space component $|\mathbf{H}^{\parallel}|$ of the diffraction vector. The short distance is S = 2.5 Å, all structure factors within $0 \leq |\mathbf{H}| \leq 2.5$ Å^{-1} have been calculated and normalized to $F(00) = 1$.

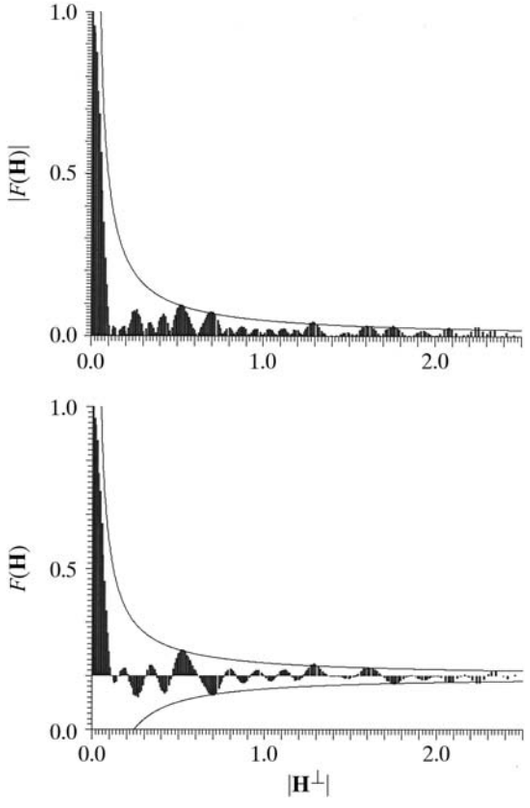


Fig. 4.6.3.9. The structure factors $F(\mathbf{H})$ (below) and their magnitudes $|F(\mathbf{H})|$ (above) of the squared Fibonacci chain decorated with equal point atoms are shown as a function of the perpendicular-space component $|\mathbf{H}^{\perp}|$ of the diffraction vector. The short distance is S = 2.5 Å, all structure factors within $0 \leq |\mathbf{H}| \leq 2.5$ Å^{-1} have been calculated and normalized to $F(00) = 1$.

are discussed. The intensities $I(\mathbf{H})$ of the Fibonacci chain decorated with point atoms are only a function of the perpendicular-space component of the diffraction vector. $|F(\mathbf{H})|$ and $F(\mathbf{H})$ are illustrated in Figs. 4.6.3.5 and 4.6.3.6 as a function of $\mathbf{H}^{\parallel}$ and of $\mathbf{H}^{\perp}$. The distribution of $|F(\mathbf{H})|$ as a function of their frequencies clearly resembles a centric distribution, as can be expected from the centrosymmetric 2D subunit cell. The shape of the distribution function depends on the radius $H_{\max}$ of the limiting sphere in reciprocal space. The number of weak reflections increases with the square of $H_{\max}$, that of strong reflections only linearly (strong reflections always have small $\mathbf{H}^{\perp}$ components).

The weighted reciprocal space of the Fibonacci sequence contains an infinite number of Bragg reflections within a limited region of the physical space. Contrary to the diffraction pattern of a periodic structure consisting of point atoms on the lattice nodes, the Bragg reflections show intensities depending on the perpendicular-space components of their diffraction vectors.

The reciprocal space of a sequence generated from hyperatoms with fractally shaped atomic surfaces (squared Fibonacci sequence) is very similar to that of the Fibonacci sequence (Figs. 4.6.3.8 and 4.6.3.9). However, there are significantly more weak reflections in the diffraction pattern of the 'fractal' sequence, caused by the geometric form factor.

4.6.3.3.1.5. Relationships between structure factors at symmetry-related points of the Fourier image

The two possible point-symmetry groups in the 1D quasiperiodic case, $K^{1D} = 1$ and $K^{1D} = \bar{1}$, relate the structure factors to

$$1: \qquad F(\mathbf{H}) = -F(\bar{\mathbf{H}}),$$
$$\bar{1}: \qquad F(\mathbf{H}) = F(\bar{\mathbf{H}}).$$

A 3D structure with 1D quasiperiodicity results from the stacking of atomic layers with distances following a quasiperiodic sequence. The point groups K^{3D} describing the symmetry of such structures result from the direct product $K^{3D} = K^{2D} \otimes K^{1D}$. K^{2D} corresponds to one of the ten crystallographic 2D point groups, K^{1D} can be $\{1\}$ or $\{1, m\}$. Consequently, 18 3D point groups are possible.

Since 1D quasiperiodic sequences can be described generically as incommensurately modulated structures, their possible point and space groups are equivalent to a subset of the (3 + 1)D superspace groups for IMSs with satellite vectors of the type (00γ), *i.e.* $\mathbf{q} = \gamma\mathbf{c}^*$, for the quasiperiodic direction [001] (Janssen *et al.*, 2004).

From the scaling properties of the Fibonacci sequence, some relationships between structure factors can be derived. Scaling the physical-space structure by a factor τ^n, $n \in \mathbb{Z}$, corresponds to a scaling of the perpendicular space by the inverse factor $(-\tau)^{-n}$. For the scaling of the corresponding reciprocal subspaces, the inverse factors compared to the direct spaces have to be applied.

The set of vectors $\mathbf{r}$, defining the vertices of a Fibonacci sequence $s(\mathbf{r})$, multiplied by a factor τ coincides with a subset of the vectors defining the vertices of the original sequence (Fig. 4.6.3.10). The residual vertices correspond to a particular decoration of the scaled sequence, *i.e.* the sequence $\tau^2 s(\mathbf{r})$. The Fourier transform of the sequence $s(\mathbf{r})$ then can be written as the sum of the Fourier transforms of the sequences $\tau s(\mathbf{r})$ and $\tau^2 s(\mathbf{r})$;

$$\sum_k \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_k) = \sum_k \exp(2\pi i \mathbf{H} \tau \mathbf{r}_k) + \sum_k \exp[2\pi i \mathbf{H}(\tau^2 \mathbf{r}_k + \tau)].$$

In terms of structure factors, this can be reformulated as

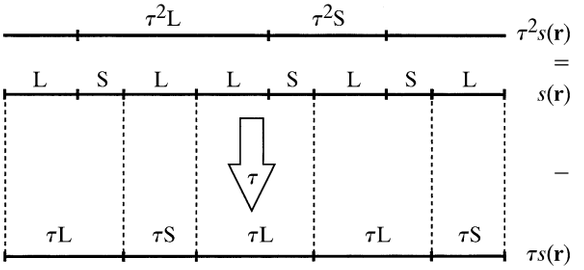


Fig. 4.6.3.10. Part ... LSLLSLSL ... of a Fibonacci sequence $s(\mathbf{r})$ before and after scaling by the factor τ. L is mapped onto τL, S onto τS = L. The vertices of the new sequence are a subset of those of the original sequence (the correspondence is indicated by dashed lines). The residual vertices $\tau^2 s(\mathbf{r})$, which give when decorating $\tau s(\mathbf{r})$ the Fibonacci sequence $s(\mathbf{r})$, form a Fibonacci sequence scaled by a factor τ^2.

$$F(\mathbf{H}) = F(\tau\mathbf{H}) + \exp(2\pi i\tau\mathbf{H})F(\tau^2\mathbf{H}).$$

Hence, phases of structure factors that are related by scaling symmetry can be determined from each other.

Further scaling relationships in reciprocal space exist: scaling a diffraction vector

$$\mathbf{H} = h_1\mathbf{d}_1^* + h_2\mathbf{d}_2^* = h_1 a^* \begin{pmatrix} 1 \\ -\tau \end{pmatrix}_V + h_2 a^* \begin{pmatrix} \tau \\ 1 \end{pmatrix}_V$$

with the matrix

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}_D,$$

$$\begin{aligned} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}_D \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}_D &= \begin{pmatrix} F_n & F_{n+1} \\ F_{n+1} & F_{n+2} \end{pmatrix}_D \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}_D \\ &= \begin{pmatrix} F_n h_1 + F_{n+1} h_2 \\ F_{n+1} h_1 + F_{n+2} h_2 \end{pmatrix}_D, \end{aligned}$$

increases the magnitudes of structure factors assigned to this particular diffraction vector $\mathbf{H}$,

$$\left|F(S^n\mathbf{H})\right| > \left|F(S^{n-1}\mathbf{H})\right| > \ldots > \left|F(S\mathbf{H})\right| > \left|F(\mathbf{H})\right|.$$

This is due to the shrinking of the perpendicular-space component of the diffraction vector by powers of $(-\tau)^{-n}$ while expanding the parallel-space component by τ^n according to the eigenvalues τ and $-\tau^{-1}$ of S acting in the two eigenspaces $\mathbf{V}^{\parallel}$ and $\mathbf{V}^{\perp}$:

$$\begin{aligned} \pi^{\parallel}(S\mathbf{H}) &= (h_2 + \tau(h_1 + h_2))a^* = (\tau h_1 + h_2(\tau + 1))a^* \\ &= \tau(h_1 + \tau h_2)a^*, \\ \pi^{\perp}(S\mathbf{H}) &= (-\tau h_2 + h_1 + h_2)a^* = (h_1 - h_2(\tau - 1))a^* \\ &= -(1/\tau)(-\tau h_1 + h_2)a^*, \\ \left|F(\tau^n\mathbf{H}^{\parallel})\right| &> \left|F(\tau^{n-1}\mathbf{H}^{\parallel})\right| > \ldots > \left|F(\tau\mathbf{H}^{\parallel})\right| > \left|F(\mathbf{H}^{\parallel})\right|. \end{aligned}$$

Thus, for scaling n times we obtain

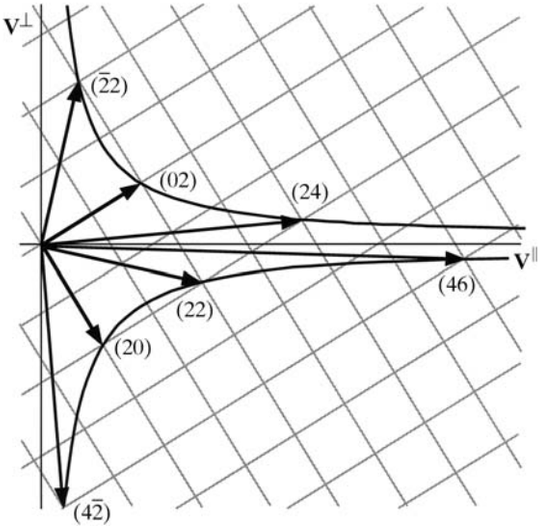


Fig. 4.6.3.11. Scaling operations of the Fibonacci sequence. The scaling operation S acts six times on the diffraction vector $\mathbf{H} = (4\bar{2})$ yielding the sequence $(4\bar{2}) \rightarrow (\bar{2}2) \rightarrow (20) \rightarrow (02) \rightarrow (22) \rightarrow (24) \rightarrow (46)$.

$$\begin{aligned} \pi^{\perp}(S^n\mathbf{H}) &= \left(-\tau\left(F_n h_1 + F_{n+1} h_2\right) + \left(F_{n+1} h_1 + F_{n+2} h_2\right)\right)a^* \\ &= \left(h_1\left(-\tau F_n + F_{n+1}\right) + h_2\left(-\tau F_{n+1} + F_{n+2}\right)\right)a^* \end{aligned}$$

with

$$\lim_{n\to\infty}\left(-\tau F_n + F_{n+1}\right) = 0 \text{ and } \lim_{n\to\infty}\left(-\tau F_{n+1} + F_{n+2}\right) = 0,$$

yielding eventually

$$\lim_{n\to\infty}\left(\pi^{\perp}(S^n\mathbf{H})\right) = 0 \text{ and } \lim_{n\to\infty}\left(F(S^n\mathbf{H})\right) = F(\mathbf{0}).$$

The scaling of the diffraction vectors $\mathbf{H}$ by S^n corresponds to a hyperbolic rotation (Janner, 1992) with angle $n\varphi$, where $\sinh\varphi = 1/2$ (Fig. 4.6.3.11):

$$\begin{aligned} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{2n} &= \begin{pmatrix} \cosh 2n\varphi & \sinh 2n\varphi \\ \sinh 2n\varphi & \cosh 2n\varphi \end{pmatrix}, \\ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{2n+1} &= \begin{pmatrix} \sinh[(2n+1)\varphi] & \cosh[(2n+1)\varphi] \\ \cosh[(2n+1)\varphi] & \sinh[(2n+1)\varphi] \end{pmatrix}. \end{aligned}$$

4.6.3.3.2. Decagonal phases

A structure quasiperiodic in two dimensions, periodic in the third dimension and with decagonal diffraction symmetry is called a decagonal phase. Its holohedral Laue symmetry group is $K = 10/mmm$. All reciprocal-space vectors $\mathbf{H} \in M^*$ can be represented on a basis (V basis) $\mathbf{a}_i^* = a_i^*(\cos 2\pi i/5, \sin 2\pi i/5, 0)$, $i = 1, \ldots, 4$ and $\mathbf{a}_5^* = a_5^*(0, 0, 1)$ (Fig. 4.6.3.12) as $\mathbf{H} = \sum_{i=1}^{5} h_i\mathbf{a}_i^*$. The vector components refer to a Cartesian coordinate system in physical (parallel) space. Thus, from the number of independent reciprocal-basis vectors necessary to index the Bragg reflections with integer numbers, the dimension of the embedding space has to be at least five. This can also be shown in a different way (Hermann, 1949).

The set M^* of all vectors $\mathbf{H}$ remains invariant under the action of the symmetry operators of the point group $10/mmm$. The symmetry-adapted matrix representations for the point-group

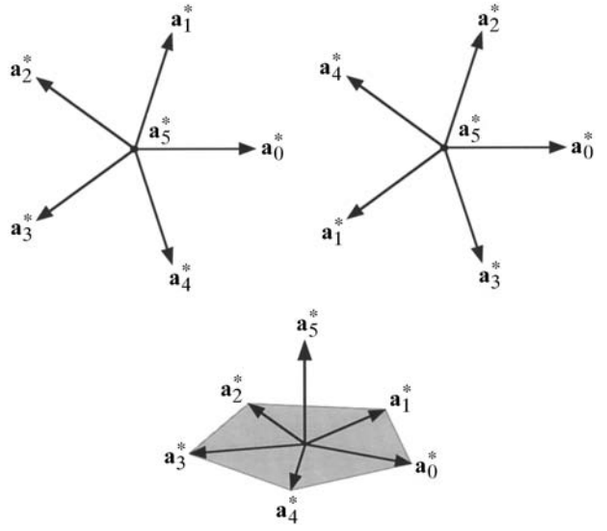


Fig. 4.6.3.12. Reciprocal basis of the decagonal phase in the 5D description projected upon $\mathbf{V}^{\parallel}$ (above left) and $\mathbf{V}^{\perp}$ (above right). Below, a perspective physical-space view is shown.

generators, the tenfold rotation $\alpha = 10$, the reflection plane $\beta = m_2$ (normal of the reflection plane along the vectors $\mathbf{a}_i^*$–$\mathbf{a}_{i+3}^*$ with $i = 1, \ldots, 4$ modulo 5) and the inversion operation $\Gamma(\gamma) = \bar{1}$ may be written in the form

$$\Gamma(\alpha) = \begin{pmatrix} 0 & 1 & \bar{1} & 0 & 0 \\ 0 & 1 & 0 & \bar{1} & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \bar{1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_D \quad \Gamma(\beta) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_D$$

$$\Gamma(\gamma) = \begin{pmatrix} \bar{1} & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & 0 & 0 \\ 0 & 0 & \bar{1} & 0 & 0 \\ 0 & 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 0 & 0 & \bar{1} \end{pmatrix}_D .$$

By block-diagonalization, these reducible symmetry matrices can be decomposed into non-equivalent irreducible representations. These can be assigned to the two orthogonal subspaces forming the 5D embedding space $\mathbf{V} = \mathbf{V}^{\parallel} \oplus \mathbf{V}^{\perp}$, the 3D parallel (physical) subspace $\mathbf{V}^{\parallel}$ and the perpendicular 2D subspace $\mathbf{V}^{\perp}$. Thus, using $W\Gamma W^{-1} = \Gamma_V = \Gamma_V^{\parallel} \oplus \Gamma_V^{\perp}$, we obtain

$$\Gamma_V(\alpha) = \left(\begin{array}{ccc|cc} \cos(\pi/5) & -\sin(\pi/5) & 0 & 0 & 0 \\ \sin(\pi/5) & \cos(\pi/5) & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & \cos(3\pi/5) & -\sin(3\pi/5) \\ 0 & 0 & 0 & \sin(3\pi/5) & \cos(3\pi/5) \end{array}\right)_V$$

$$= \left(\begin{array}{c|c} \Gamma^{\parallel}(\alpha) & 0 \\ \hline 0 & \Gamma^{\perp}(\alpha) \end{array}\right)_V ,$$

$$\Gamma_V(\beta) = \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \bar{1} \end{array}\right)_V , \quad \Gamma_V(\gamma) \left(\begin{array}{ccc|cc} \bar{1} & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & 0 & 0 \\ 0 & 0 & \bar{1} & 0 & 0 \\ \hline 0 & 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 0 & 0 & \bar{1} \end{array}\right)_V ,$$

where

$$W = \left(\begin{array}{ccccc} a_1^* \cos(2\pi/5) & a_2^* \cos(4\pi/5) & a_3^* \cos(6\pi/5) & a_4^* \cos(8\pi/5) & 0 \\ a_1^* \sin(2\pi/5) & a_2^* \sin(4\pi/5) & a_3^* \sin(6\pi/5) & a_4^* \sin(8\pi/5) & 0 \\ 0 & 0 & 0 & 0 & a_5^* \\ \hline a_1^* \cos(6\pi/5) & a_2^* \cos(2\pi/5) & a_3^* \cos(8\pi/5) & a_4^* \cos(4\pi/5) & 0 \\ a_1^* \sin(6\pi/5) & a_2^* \sin(2\pi/5) & a_3^* \sin(8\pi/5) & a_4^* \sin(4\pi/5) & 0 \end{array}\right).$$

The column vectors of the matrix W give the parallel- (above the partition line) and perpendicular-space components (below the partition line) of a reciprocal basis in V space. Thus, W can be rewritten using the physical-space reciprocal basis defined above as

$$W = \left(\mathbf{d}_1^*, \quad \mathbf{d}_2^*, \quad \mathbf{d}_3^*, \quad \mathbf{d}_4^*, \quad \mathbf{d}_5^*\right),$$

yielding the reciprocal basis $\mathbf{d}_i^*, i = 1, \ldots, 5$, in the 5D embedding space (D space):

$$\mathbf{d}_i^* = a_i^* \begin{pmatrix} \cos(2\pi i/5) \\ \sin(2\pi i/5) \\ 0 \\ \cos(6\pi i/5) \\ \sin(6\pi i/5) \end{pmatrix}_V , \; i = 1, \ldots, 4 \text{ and } \mathbf{d}_5^* = a_5^* \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}_V .$$

The 5×5 symmetry matrices can each be decomposed into a 3×3 matrix and a 2×2 matrix. The first one, $\Gamma^{\parallel}$, acts on the parallel-space component, the second one, $\Gamma^{\perp}$, on the perpendicular-space component. In the case of $\Gamma(\alpha)$, the coupling factor between a rotation in parallel and perpendicular space is 3. Thus, a $\pi/5$ rotation in physical space is related to a $3\pi/5$ rotation in perpendicular space (Fig. 4.6.3.12).

With the condition $\mathbf{d}_i \cdot \mathbf{d}_j^* = \delta_{ij}$, a basis in direct 5D space is obtained:

$$\mathbf{d}_i = \frac{2}{5a_i^*} \begin{pmatrix} \cos(2\pi i/5) - 1 \\ \sin(2\pi i/5) \\ 0 \\ \cos(6\pi i/5) - 1 \\ \sin(6\pi i/5) \end{pmatrix}, \; i = 1, \ldots, 4, \text{ and } \mathbf{d}_5 = \frac{1}{a_5^*} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

The metric tensors G, G^* are of the type

$$\begin{pmatrix} A & C & C & C & 0 \\ C & A & C & C & 0 \\ C & C & A & C & 0 \\ C & C & C & A & 0 \\ 0 & 0 & 0 & 0 & B \end{pmatrix}$$

with $A = 2a_1^{*2}$, $B = a_5^{*2}$, $C = -(1/2)a_1^{*2}$ for the reciprocal space and $A = 4/5a_1^{*2}$, $B = 1/a_5^{*2}$, $C = 2/5a_1^{*2}$ for the direct space. Thus, for the lattice parameters in reciprocal space we obtain $d_i^* = a_i^*(2)^{1/2}$, $i = 1, \ldots, 4$; $d_5^* = a_5^*$; $\alpha_{ij}^* = 104.5^\circ$, $i, j = 1, \ldots, 4$; $\alpha_{i5}^* = 90^\circ$, $i = 1, \ldots, 4$, and for those in direct space $d_i = 2/[a_i^*(5)^{1/2}]$, $i = 1, \ldots, 4$; $d_5 = 1/a_5^*$; $\alpha_{ij} = 60^\circ$, $i, j = 1, \ldots, 4$; $\alpha_{i5} = 90^\circ$, $i = 1, \ldots, 4$. The volume of the 5D unit cell can be calculated from the metric tensor G:

$$V = [\det(G)]^{1/2} = \frac{4}{5(5)^{1/2}(a_1^*)^4 a_5^*} = \frac{(5)^{1/2}}{4}(d_1)^4 d_5.$$

Since decagonal phases are only quasiperiodic in two dimensions, it is sufficient to demonstrate their characteristics on a 2D

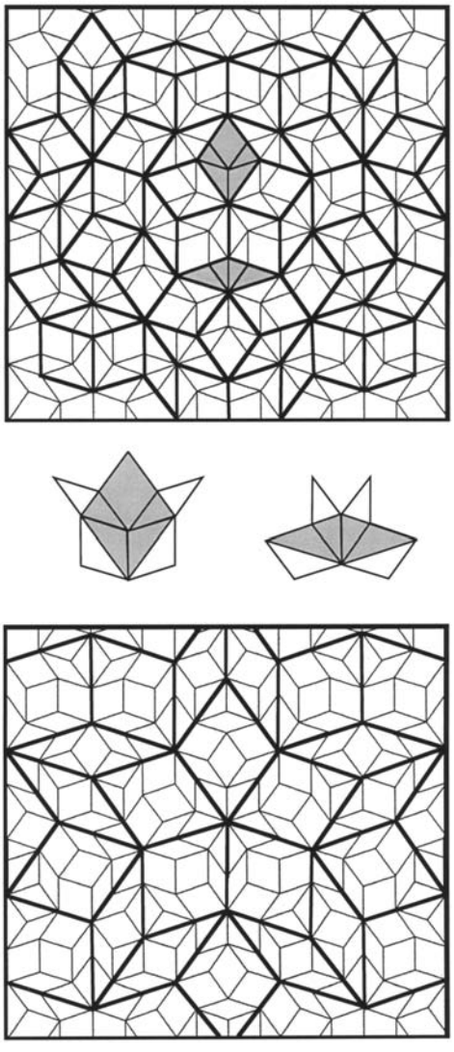


Fig. 4.6.3.13. A section of a Penrose tiling (thin lines) superposed by its τ-deflated tiling (above, thick lines) and by its τ^2-deflated tiling (below, thick lines). In the middle, the inflation rule of the Penrose tiling is illustrated.

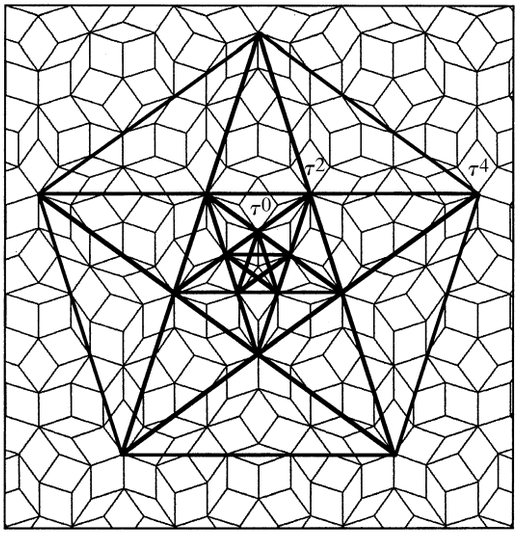


Fig. 4.6.3.14. Scaling symmetry of a pentagram superposed on the Penrose tiling. A vector pointing to a corner of a pentagon (star) is mapped by the rotoscaling operation (rotation around $\pi/5$ and dilatation by a factor τ^2) onto the next largest pentagon (star).

example, the canonical *Penrose tiling* (Penrose, 1974). It can be constructed from two unit tiles: a skinny (acute angle $\alpha_s = \pi/5$) and a fat (acute angle $\alpha_f = 2\pi/5$) rhomb with equal edge lengths a_r and areas $A_S = a_r^2 \sin(\pi/5)$, $A_F = a_r^2 \sin(2\pi/5)$ (Fig. 4.6.3.13). The areas and frequencies of these two unit tiles in the Penrose tiling are both in a ratio 1 to τ. By replacing each skinny and fat rhomb according to the inflation rule, a τ-inflated tiling is obtained. Inflation (deflation) means that the number of tiles is inflated (deflated), their edge lengths are decreased (increased) by a factor τ. By infinite repetition of this inflation operation, an infinite Penrose tiling is generated. Consequently, this substitution operation leaves the tiling invariant.

From Fig. 4.6.3.13 it can be seen that the sets of vertices of the deflated tilings are subsets of the set of vertices of the original tiling. The τ-deflated tiling is dual to the original tiling; a further deflation by a factor τ gives the original tiling again. However, the edge lengths of the tiles are increased by a factor τ^2, and the tiling is rotated around 36°. Only the fourth deflation of the original tiling yields the original tiling in its original orientation but with all lengths multiplied by a factor τ^4.

Contrary to the reciprocal-space scaling behaviour of $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{4} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$, the set of vertices $M = \{\mathbf{r} = \sum_{i=1}^{4} n_i \mathbf{a}_i | n_i \in \mathbb{Z}\}$ of the Penrose tiling is not invariant by scaling the length scale simply by a factor τ using the scaling matrix S:

$$S = \begin{pmatrix} 0 & 1 & 0 & \bar{1} \\ 0 & 1 & 1 & \bar{1} \\ \bar{1} & 1 & 1 & 0 \\ \bar{1} & 0 & 1 & 0 \end{pmatrix}_D \text{ acting on vectors } \mathbf{r} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \end{pmatrix}_D .$$

The square of S, however, maps all vertices of the Penrose tiling upon other ones:

$$S^2 = \begin{pmatrix} 1 & 1 & 0 & \bar{1} \\ 0 & 2 & 1 & \bar{1} \\ \bar{1} & 1 & 2 & 0 \\ \bar{1} & 0 & 1 & 1 \end{pmatrix}_D , \quad \Gamma(\alpha)S^2 = \begin{pmatrix} 1 & 1 & \bar{1} & \bar{1} \\ 1 & 2 & 0 & \bar{2} \\ 0 & 2 & 1 & \bar{1} \\ \bar{1} & 1 & 1 & 0 \end{pmatrix}_D .$$

S^2 corresponds to a hyperbolic rotation with $\chi = \cosh^{-1}(3/2)$ in superspace (Janner, 1992). However, only operations of the type S^{4n}, $n = 0, 1, 2 \ldots$, scale the Penrose tiling in a way which is equivalent to the ($4n$th) substitutional operations discussed above. The rotoscaling operation $\Gamma(\alpha)S^2$, also a symmetry operation of the Penrose tiling, leaves a pentagram invariant as demonstrated in Fig. 4.6.3.14 (Janner, 1992). Block-diagonalization of the scaling matrix S decomposes it into two non-equivalent irreducible representations which give the scaling properties in the two orthogonal subspaces of the 4D embedding space, $\mathbf{V} = \mathbf{V}^{\parallel} \oplus \mathbf{V}^{\perp}$, the 2D parallel (physical) subspace $\mathbf{V}^{\parallel}$ and the perpendicular 2D subspace $\mathbf{V}^{\perp}$. Thus, using $WSW^{-1} = S_V = S_V^{\parallel} \oplus S_V^{\perp}$, we obtain

$$S_V = \left(\begin{array}{cc|cc} \tau & 0 & 0 & 0 \\ 0 & \tau & 0 & 0 \\ \hline 0 & 0 & -1/\tau & 0 \\ 0 & 0 & 0 & -1/\tau \end{array}\right)_V = \left(\begin{array}{cc|cc} S_V^{\parallel} & & 0 & \\ & & & \\ \hline 0 & & S_V^{\perp} & \end{array}\right)_V ,$$

where

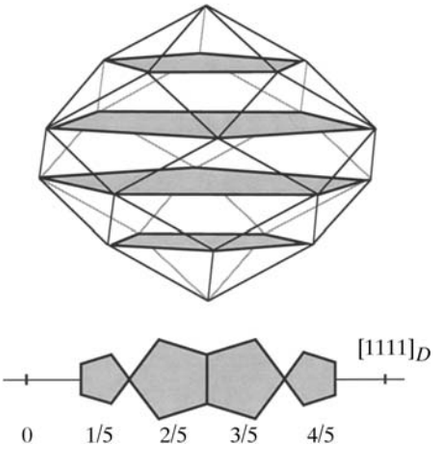


Fig. 4.6.3.15. Atomic surface of the Penrose tiling in the 5D hypercubic description. The projection of the 5D hypercubic unit cell upon $\mathbf{V}^{\perp}$ gives a rhomb-icosahedron (above). The Penrose tiling is generated by four equidistant pentagons (shaded) inscribed in the rhomb-icosahedron. Below is a perpendicular-space projection of the same pentagons, which are located on the $[1111]_D$ diagonal of the 4D hyper-rhombohedral unit cell in the 4D description.

$$W = \begin{pmatrix} a_1^* \cos(2\pi/5) & a_2^* \cos(4\pi/5) & a_3^* \cos(6\pi/5) & a_4^* \cos(8\pi/5) \\ a_1^* \sin(2\pi/5) & a_2^* \sin(4\pi/5) & a_3^* \sin(6\pi/5) & a_4^* \sin(8\pi/5) \\ \hline a_1^* \cos(4\pi/5) & a_2^* \cos(8\pi/5) & a_3^* \cos(2\pi/5) & a_4^* \cos(6\pi/5) \\ a_1^* \sin(4\pi/5) & a_2^* \sin(8\pi/5) & a_3^* \sin(2\pi/5) & a_4^* \sin(6\pi/5) \end{pmatrix}.$$

The 2D Penrose tiling can also be embedded canonically in the 5D space. Canonically means that the 5D lattice is hypercubic and that the projection of one unit cell upon the 3D perpendicular space $\mathbf{V}^{\perp}$, giving a rhomb-icosahedron, defines the atomic surface. However, the parallel-space image $\mathbf{a}_i^*$, $i = 1, \ldots, 4$, with $\mathbf{a}_0^* = -(\mathbf{a}_1^* + \mathbf{a}_2^* + \mathbf{a}_3^* + \mathbf{a}_4^*)$, of the 5D basis $\mathbf{d}_i^*$, $i = 1, \ldots, 4$ is not linearly independent. Consequently, the atomic surface consists of only a subset of the points contained in the rhomb-icosahedron: five equidistant pentagons (one with diameter zero) resulting as sections of the rhomb-icosahedron with five equidistant parallel planes (Fig. 4.6.3.15). The linear dependence of the 5D basis allows the embedding in the 4D space. The resulting hyper-rhombohedral hyperlattice is spanned by the basis $\mathbf{d}_i$, $i = 1, \ldots, 4$, discussed above. The atomic surfaces occupy the positions $p/5(1111)$, $p = 1, \ldots, 4$, on the body diagonal of the 4D unit cell. Neighbouring pentagons are in an *anti* position to each other (Fig. 4.6.3.16). Thus the 4D unit cell is decorated centrosymmetrically. The edge length a_r of a Penrose rhomb is related to the length of physical-space basis vectors a_i^* by $a_r = \tau S$, with the smallest distance $S = (2\tau/5a_i^*)$, $i = 1, \ldots, 4$. The *point density* (number of vertices per unit area) of a Penrose tiling with Penrose rhombs of edge length a_r can be calculated from the ratio of the relative number of unit tiles in the tiling to their area:

$$\rho = \frac{1+\tau}{a_r^2[\sin(\pi/5) + \tau \sin(2\pi/5)]} = (5/2)a_i^{*2}(2-\tau)^2 \tan(2\pi/5).$$

This is equivalent to the calculation from the 4D description,

$$\rho = \frac{\sum_{i=1}^{4} \Omega_{\rm AS}^i}{\Omega_{\rm UC}} = \frac{\sum_{i=1}^{4} (5/2)\lambda^2 \sin(2\pi/5)}{4/[5(5)^{1/2}|a_i^*|^4]} = (5/2)a_i^{*2}(2-\tau)^2 \tan(2\pi/5),$$

where $\Omega_{\rm AS}$ and $\Omega_{\rm UC}$ are the area of the atomic surface and the volume of the 4D unit cell, respectively. The pentagon radii are

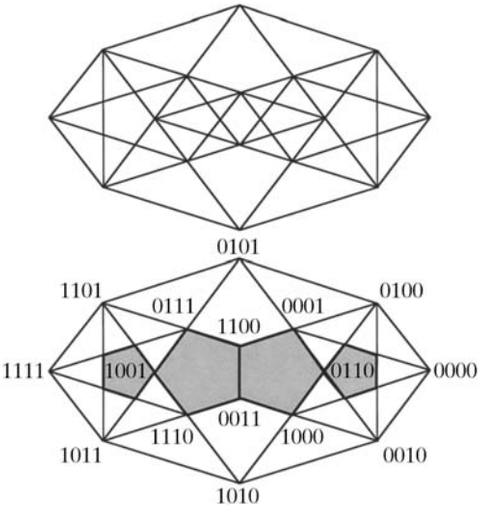


Fig. 4.6.3.16. Projection of the 4D hyper-rhombohedral unit cell of the Penrose tiling in the 4D description upon the perpendicular space. In the upper drawing all edges between the 16 corners are shown. In the lower drawing the corners are indexed and the four pentagonal atomic surfaces of the Penrose tiling are shaded.

$\lambda_{1,4} = 2(2-\tau)/5a^*$ and $\lambda_{2,3} = 2(\tau-1)/5a^*$ for the atomic surfaces in $(p/5)(1111)$ with $p = 1, 4$ and $p = 2, 3$. A detailed discussion of the properties of Penrose tiling is given in the papers of Penrose (1974, 1979), Jaric (1986) and Pavlovitch & Kleman (1987).

4.6.3.3.2.1. Indexing

The indexing of the submodule M_1^* of the diffraction pattern of a decagonal phase is not unique. Since M_1^* corresponds to a $\mathbb{Z}$ module of rank 4 with decagonal point symmetry, it is invariant under scaling by $\tau^n, n \in \mathbb{Z}$: $S^n M^* = \tau^n M^*$. Nevertheless, an optimum basis (low indices are assigned to strong reflections) can be derived: not the metrics, as for regular periodic crystals, but the intensity distribution characterizes the best choice of indexing.

A correct set of reciprocal-basis vectors can be identified experimentally in the following way:

(1) Find directions of systematic absences or pseudo-absences determining the possible orientations of the reciprocal-basis vectors (see Rabson *et al.*, 1991).

(2) Find pairs of strong reflections whose physical-space diffraction vectors are related to each other by the factor τ.

(3) Index these reflections by assigning an appropriate value to a^*. This value should be derived from the shortest interatomic distance S and the edge length of the unit tiles expected in the structure.

(4) The reciprocal basis is correct if all observable Bragg reflections can be indexed with integer numbers.

4.6.3.3.2.2. Diffraction symmetry

The diffraction symmetry of decagonal phases can be described by the Laue groups $10/mmm$ or $10/m$. The set of all vectors $\mathbf{H}$ forms a Fourier module $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{5} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ of rank 5 in physical space which can be decomposed into two submodules $M^* = M_1^* \oplus M_2^*$. $M_1^* = \{h_1\mathbf{a}_1^* + h_2\mathbf{a}_2^* + h_3\mathbf{a}_3^* + h_4\mathbf{a}_4^*\}$ corresponds to a $\mathbb{Z}$ module of rank 4 in a 2D subspace, $M_2^* = \{h_5\mathbf{a}_5^*\}$ corresponds to a $\mathbb{Z}$ module of rank 1 in a 1D subspace. Consequently, the first submodule can be considered as a projection from a 4D reciprocal lattice, $M_1^* = \pi^{\parallel}(\Sigma^*)$, while the second submodule is of the form of a regular 1D reciprocal lattice, $M_2^* = \Lambda^*$. The diffraction pattern of the Penrose tiling decorated with equal point scatterers on its vertices is shown in Fig. 4.6.3.17. All Bragg reflections within

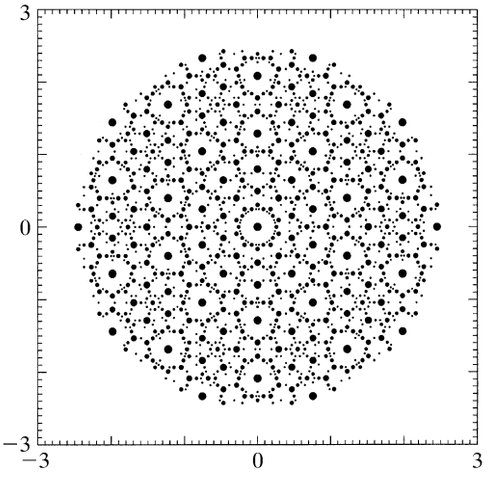


Fig. 4.6.3.17. Schematic diffraction pattern of the Penrose tiling (edge length of the Penrose unit rhombs a_r = 4.04 Å). All reflections are shown within $10^{-2}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $0 \le |\mathbf{H}^{\parallel}| \le 2.5$ Å^{-1}.

$10^{-2}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ are depicted. Without intensity-truncation limit, the diffraction pattern would be densely filled with discrete Bragg reflections. To illustrate their spatial and intensity distribution, an enlarged section of Fig. 4.6.3.17 is shown in Fig. 4.6.3.18. This picture shows all Bragg reflections within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$. The projected 4D reciprocal-lattice unit cell is drawn and several reflections are indexed. All reflections are arranged along lines in five symmetry-equivalent orientations. The perpendicular-space diffraction patterns (Figs. 4.6.3.19 and 4.6.3.20) show a characteristic star-like distribution of the Bragg reflections. This is a consequence of the pentagonal shape of the atomic surfaces: the Fourier transform of a pentagon has a star-like distribution of strong Fourier coefficients.

The 5D decagonal space groups that may be of relevance for the description of decagonal phases are listed in Table 4.6.3.1. These space groups are a subset of all 5D decagonal space groups fulfilling the condition that the 5D point groups they are associated with are isomorphous to the 3D point groups describing the diffraction symmetry. Their structures are comparable to 3D hexagonal groups. Hence, only primitive lattices exist. The orientation of the symmetry elements in the 5D space is defined by the isomorphism of the 3D and 5D point groups. However, the action of the tenfold rotation is different in the subspaces $\mathbf{V}^{\parallel}$ and

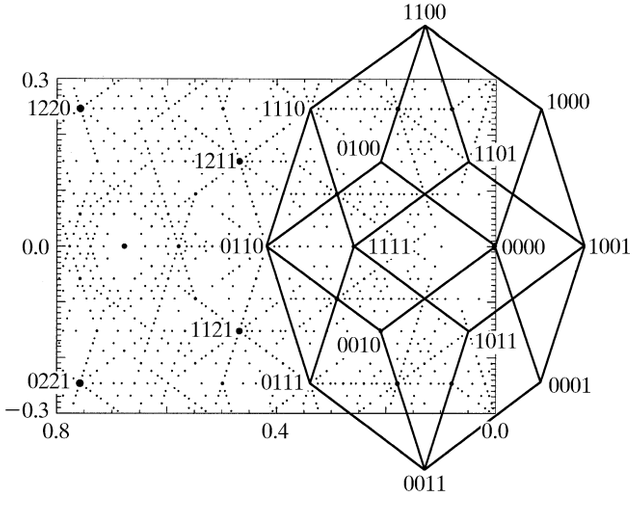


Fig. 4.6.3.18. Enlarged section of Fig. 4.6.3.17. All reflections shown are selected within the given limits from a data set within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $0 \le |\mathbf{H}^{\parallel}| \le 2.5$ Å^{-1}. The projected 4D reciprocal-lattice unit cell is drawn and several reflections are indexed.

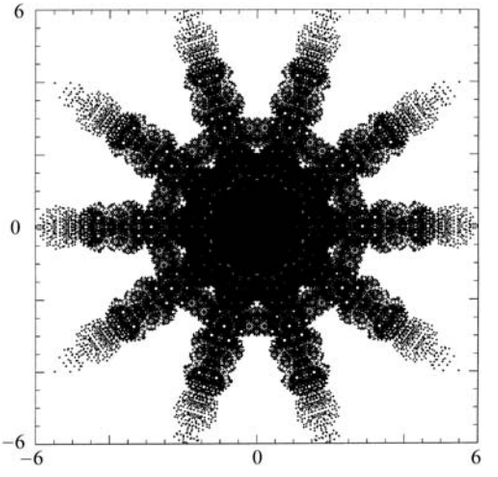


Fig. 4.6.3.19. The perpendicular-space diffraction pattern of the Penrose tiling (edge length of the Penrose unit rhombs a_r = 4.04 Å). All reflections are shown within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $0 \le |\mathbf{H}^{\parallel}| \le 2.5$ Å^{-1}.

$\mathbf{V}^{\perp}$: a rotation of $\pi/5$ in $\mathbf{V}^{\parallel}$ is correlated with a rotation of $3\pi/5$ in $\mathbf{V}^{\perp}$. The reflection and inversion operations are equivalent in both subspaces.

4.6.3.3.2.3. Structure factor

The structure factor for the decagonal phase corresponds to the Fourier transform of the 5D unit cell,

$$F(\mathbf{H}) = \sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}, \mathbf{H}^{\perp}) g_k(\mathbf{H}^{\perp}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_k),$$

with 5D diffraction vectors $\mathbf{H} = \sum_{i=1}^{5} h_i \mathbf{d}_i^*$, N hyperatoms, parallel-space atomic scattering factor $f_k(\mathbf{H}^{\parallel})$, temperature factor $T_k(\mathbf{H}^{\parallel}, \mathbf{H}^{\perp})$ and perpendicular-space geometric form factor $g_k(\mathbf{H}^{\perp})$. $T_k(\mathbf{H}^{\parallel}, \mathbf{0})$ is equivalent to the conventional Debye–Waller factor and $T_k(\mathbf{0}, \mathbf{H}^{\perp})$ describes random fluctuations along the perpendicular-space coordinate. These fluctuations cause characteristic jumps of vertices in physical space (*phason flips*). Even random phason flips map the vertices onto positions which can still be described by physical-space vectors of the type $\mathbf{r} = \sum_{i=1}^{5} n_i \mathbf{a}_i$. Consequently, the set $M = \{\mathbf{r} = \sum_{i=1}^{5} n_i \mathbf{a}_i | n_i \in \mathbb{Z}\}$ of all possible vectors forms a $\mathbb{Z}$ module. The shape of the atomic surfaces corresponds to a selection rule for the positions actually occupied. The geometric form factor $g_k(\mathbf{H}^{\perp})$ is equivalent to the

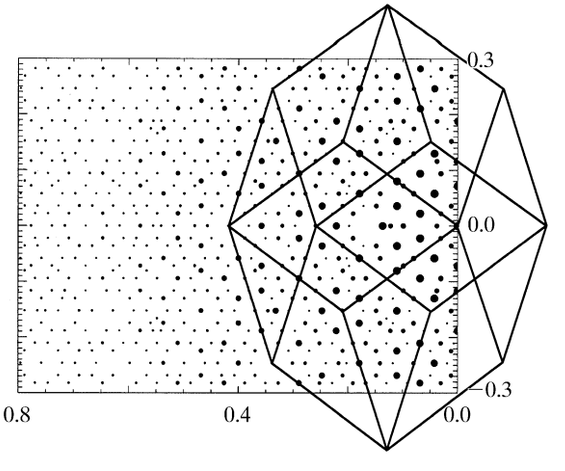


Fig. 4.6.3.20. Enlarged section of Fig. 4.6.3.19 showing the projected 4D reciprocal-lattice unit cell.

Table 4.6.3.1. *3D point groups of order k describing the diffraction symmetry and corresponding 5D decagonal space groups with reflection conditions (see Rabson et al., 1991)*

3D point group	k	5D space group	Reflection condition
$\frac{10}{m}\frac{2}{m}\frac{2}{m}$	40	$P\frac{10}{m}\frac{2}{m}\frac{2}{m}$	No condition
		$P\frac{10}{m}\frac{2}{c}\frac{2}{c}$	$h_1h_2h_2h_1h_5$: $h_5 = 2n$ $h_1h_2\overline{h_2h_1}h_5$: $h_5 = 2n$
		$P\frac{10_5}{m}\frac{2}{m}\frac{2}{c}$	$h_1h_2\overline{h_2h_1}h_5$: $h_5 = 2n$
		$P\frac{10_5}{m}\frac{2}{c}\frac{2}{m}$	$h_1h_2h_2h_1h_5$: $h_5 = 2n$
$\frac{10}{m}$	20	$P\frac{10}{m}$	No condition
		$P\frac{10_5}{m}$	$0000h_5$: $h_5 = 2n$
1022	20	$P1022$	No condition
		$P10_j22$	$0000h_5$: $jh_5 = 10n$
$10mm$	20	$P10mm$	No condition
		$P10cc$	$h_1h_2h_2h_1h_5$: $h_5 = 2n$ $h_1h_2\overline{h_2h_1}h_5$: $h_5 = 2n$
		$P10_5mc$	$h_1h_2\overline{h_2h_1}h_5$: $h_5 = 2n$
		$P10_5cm$	$h_1h_2h_2h_1h_5$: $h_5 = 2n$
$\overline{1}0m2$	20	$P\overline{1}0m2$	No condition
		$P\overline{1}0c2$	$h_1h_2h_2h_1h_5$: $h_5 = 2n$
		$P\overline{1}02m$	No condition
		$P\overline{1}02c$	$h_1h_2\overline{h_2h_1}h_5$: $h_5 = 2n$
$\overline{10}$	10	$P\overline{10}$	No condition
10	10	$P10$	No condition
		$P10_j$	$0000h_5$: $jh_5 = 10n$

Fourier transform of the *atomic surface*, *i.e.* the 2D perpendicular-space component of the 5D *hyperatoms*.

For example, the canonical Penrose tiling $g_k(\mathbf{H}^\perp)$ corresponds to the Fourier transform of pentagonal atomic surfaces:

$$g_k(\mathbf{H}^\perp) = (1/A_{\rm UC}^\perp)\int_{A_k} \exp(2\pi i\mathbf{H}^\perp\cdot\mathbf{r})\ \mathrm{d}\mathbf{r},$$

where $A_{\rm UC}^\perp$ is the area of the 5D unit cell projected upon $\mathbf{V}^\perp$ and A_k is the area of the kth atomic surface. The area $A_{\rm UC}^\perp$ can be calculated using the formula

$$A_{\rm UC}^\perp = (4/25a_i^{*2})[(7+\tau)\sin(2\pi/5) + (2+\tau)\sin(4\pi/5)].$$

Evaluating the integral by decomposing the pentagons into triangles, one obtains

$$g_k(\mathbf{H}^\perp) = \frac{1}{A_{\rm UC}^\perp}\sin\left(\frac{2\pi}{5}\right) \times \sum_{j=0}^{4}\frac{A_j[\exp(iA_{j+1}\lambda_k)-1] - A_{j+1}[\exp(iA_j\lambda_k)-1]}{A_jA_{j+1}(A_j - A_{j+1})}$$

with $j = 0, \ldots, 4$ running over the five triangles, where the radii of the pentagons are λ_j, $A_j = 2\pi\mathbf{H}^\perp\mathbf{e}_j$,

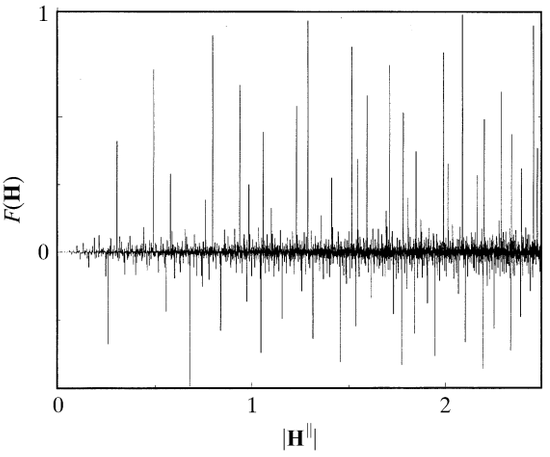


Fig. 4.6.3.21. Radial distribution function of the structure factors $F(\mathbf{H})$ of the Penrose tiling (edge length of the Penrose unit rhombs a_r = 4.04 Å) decorated with point atoms as a function of $\mathbf{H}^\parallel$. All structure factors within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $0 \le |\mathbf{H}^\parallel| \le 2.5$ Å^{-1} have been used and normalized to $F(0000) = 1$.

$$\mathbf{H}^\perp = \pi^\perp(\mathbf{H}) = \sum_{j=0}^{4} h_j a_j^* \begin{pmatrix} 0 \\ 0 \\ 0 \\ \cos(6\pi j/5) \\ \sin(6\pi j/5) \end{pmatrix}$$

and the vectors

$$\mathbf{e}_j = \frac{1}{a_j^*}\begin{pmatrix} 0 \\ 0 \\ 0 \\ \cos(2\pi j/5) \\ \sin(2\pi j/5) \end{pmatrix} \text{ with } j = 0, \ldots, 4.$$

As shown by Ishihara & Yamamoto (1988), the Penrose tiling can be considered to be a superstructure of a pentagonal tiling with only one type of pentagonal atomic surface in the nD unit cell. Thus, for the Penrose tiling, three special reflection classes can be distinguished: for $|\sum_{i=1}^4 h_i| = m \bmod 5$ and $m = 0$ the class of strong main reflections is obtained, and for $m = \pm1, \pm2$ the classes of weaker first- and second-order satellite reflections are obtained (see Fig. 4.6.3.18).

4.6.3.3.2.4. Intensity statistics

This section deals with the reciprocal-space characteristics of the 2D quasiperiodic component of the 3D structure, namely the Fourier module M_1^*. The radial structure-factor distributions of the Penrose tiling decorated with point scatterers are plotted in Figs. 4.6.3.21 and 4.6.3.22 as a function of parallel and perpendicular space. The distribution of $|F(\mathbf{H})|$ as a function of their frequencies clearly resembles a centric distribution, as can be expected from the centrosymmetric 4D subunit cell. The shape of the distribution function depends on the radius of the limiting sphere in reciprocal space. The number of weak reflections increases to the power of four, that of strong reflections only quadratically (strong reflections always have small $\mathbf{H}^\perp$ components). The radial distribution of the structure-factor amplitudes as a function of perpendicular space clearly shows three branches, corresponding to the reflection classes $\sum_{i=1}^4 h_i = m \bmod 5$ with $|m| = 0$, $|m| = 1$ and $|m| = 2$ (Fig. 4.6.3.23).

The weighted reciprocal space of the Penrose tiling contains an infinite number of Bragg reflections within a limited region of the physical space. Contrary to the diffraction pattern of a periodic

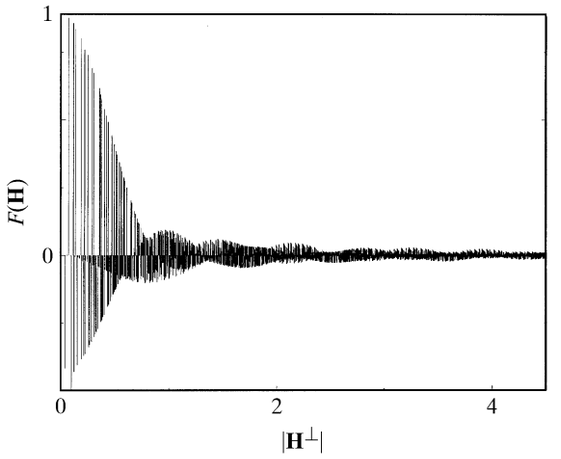


Fig. 4.6.3.22. Radial distribution function of the structure factors $F(\mathbf{H})$ of the Penrose tiling (edge length of the Penrose unit rhombs a_r = 4.04 Å) decorated with point atoms as a function of $\mathbf{H}^{\perp}$. All structure factors within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $0 \le |\mathbf{H}^{\parallel}| \le 2.5$ Å^{-1} have been used and normalized to $F(0000) = 1$.

structure consisting of point atoms on the lattice nodes, the Bragg reflections show intensities depending on the perpendicular-space components of their diffraction vectors (Figs. 4.6.3.19, 4.6.3.20 and 4.6.3.22).

4.6.3.3.2.5. Relationships between structure factors at symmetry-related points of the Fourier image

Scaling the Penrose tiling by a factor τ^{-n} by employing the matrix S^{-n} scales at the same time its reciprocal space by a factor τ^n:

$$\boldsymbol{S}\mathbf{H} = \begin{pmatrix} 0 & 1 & 0 & \bar{1} & 0 \\ 0 & 1 & 1 & \bar{1} & 0 \\ \bar{1} & 1 & 1 & 0 & 0 \\ \bar{1} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_D \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{pmatrix} = \begin{pmatrix} h_2 - h_4 \\ h_2 + h_3 - h_4 \\ -h_1 + h_2 + h_3 \\ -h_1 + h_3 \\ h_5 \end{pmatrix}.$$

Since this operation increases the lengths of the diffraction vectors by the factor τ in parallel space and decreases them by the factor $1/\tau$ in perpendicular space, the following distribution of structure-factor magnitudes (for point atoms at rest) is obtained:

$$\begin{aligned} \big|F(S^n\mathbf{H})\big| > \big|F(S^{n-1}\mathbf{H})\big| > \ldots > \big|F(S^1\mathbf{H})\big| > \big|F(\mathbf{H})\big|, \\ \big|F(\tau^n\mathbf{H}^{\parallel})\big| > \big|F(\tau^{n-1}\mathbf{H}^{\parallel})\big| > \ldots > \big|F(\tau\mathbf{H}^{\parallel})\big| > \big|F(\mathbf{H})\big|. \end{aligned}$$

The scaling operations S^n, $n \in \mathbb{Z}$, the rotoscaling operations $(\Gamma(\alpha)S^2)^n$ (Fig. 4.6.3.14) and the tenfold rotation $(\Gamma(\alpha))^n$, where

$$(\Gamma(\alpha)S^2)^n = \begin{pmatrix} 1 & 1 & \bar{1} & \bar{1} & 0 \\ 1 & 2 & 0 & \bar{2} & 0 \\ 0 & 2 & 1 & \bar{1} & 0 \\ \bar{1} & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^n_D,$$

connect all structure factors with diffraction vectors pointing to the nodes of an infinite series of pentagrams. The structure factors with positive signs are predominantly on the vertices of the pentagram while the ones with negative signs are arranged on circles around the vertices (Figs. 4.6.3.24 to 4.6.3.27).

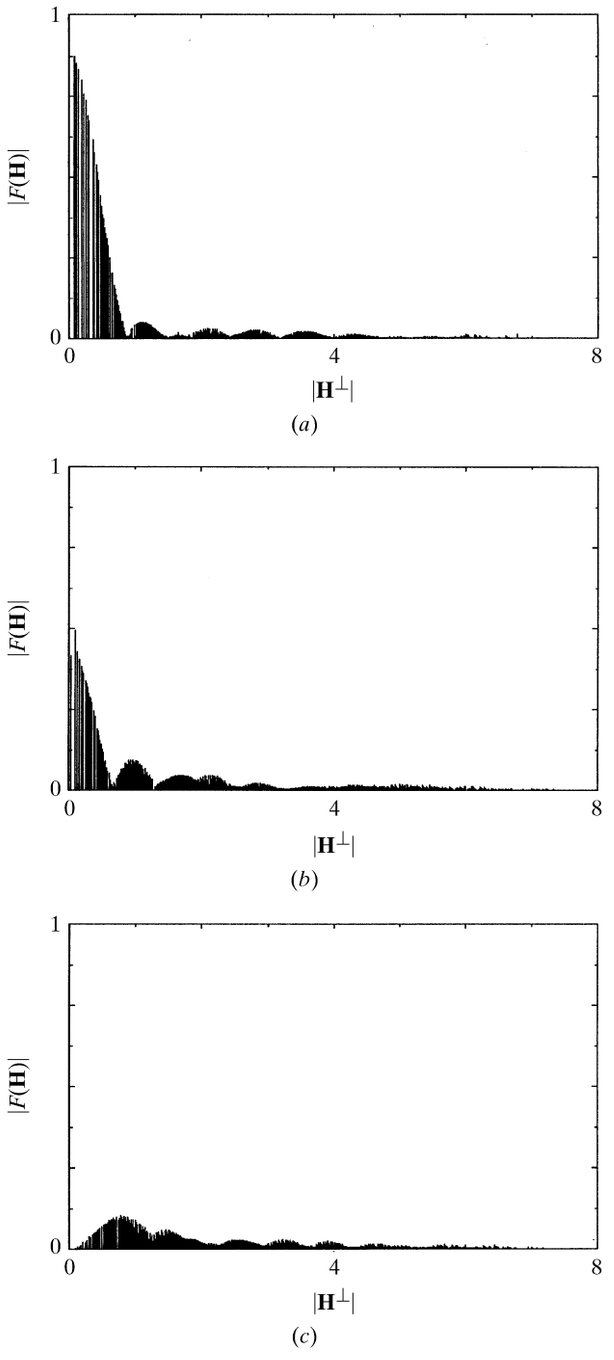


Fig. 4.6.3.23. Radial distribution function of the structure-factor magnitudes $|F(\mathbf{H})|$ of the Penrose tiling (edge length of the Penrose unit rhombs a_r = 4.04 Å) decorated with point atoms as a function of $\mathbf{H}^{\perp}$. All structure factors within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $0 \le |\mathbf{H}^{\parallel}| \le 2.5$ Å^{-1} have been used and normalized to $F(0000) = 1$. The branches with (*a*) $|\sum_{i=1}^4 h_i| = 0 \bmod 5$, (*b*) $|\sum_{i=1}^4 h_i| = 1 \bmod 5$ and (*c*) $|\sum_{i=1}^4 h_i| = 2 \bmod 5$ are shown.

4.6.3.3.3. Icosahedral phases

A structure that is quasiperiodic in three dimensions and exhibits icosahedral diffraction symmetry is called an icosahedral phase. Its holohedral Laue symmetry group is $K = m\bar{3}\bar{5}$. All reciprocal-space vectors $\mathbf{H} = \sum_{i=1}^6 h_i \mathbf{a}_i^* \in M^*$ can be represented on a basis $\mathbf{a}_1^* = a^*(0, 0, 1)$, $\mathbf{a}_i^* = a^*[\sin\theta\cos(2\pi i/5), \sin\theta\sin(2\pi i/5), \cos\theta]$, $i = 2, \ldots, 6$ where $\sin\theta = 2/(5)^{1/2}$, $\cos\theta = 1/(5)^{1/2}$ and $\theta \simeq 63.44^\circ$, the angle between two neighbouring fivefold axes (Fig. 4.6.3.28). This can be rewritten as

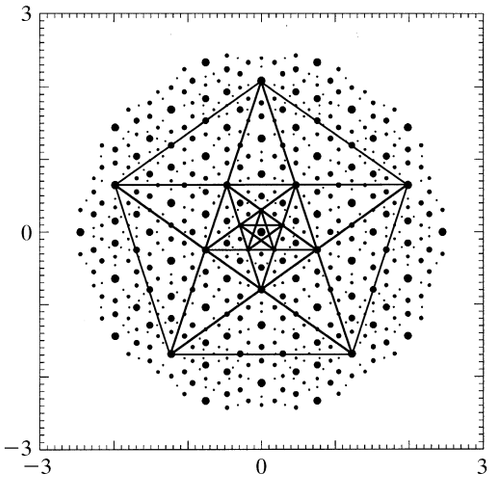


Fig. 4.6.3.24. Pentagrammal relationships between scaling symmetry-related positive structure factors $F(\mathbf{H})$ of the Penrose tiling (edge length $a_r = 4.04$ Å) in parallel space. The magnitudes of the structure factors are indicated by the diameters of the filled circles.

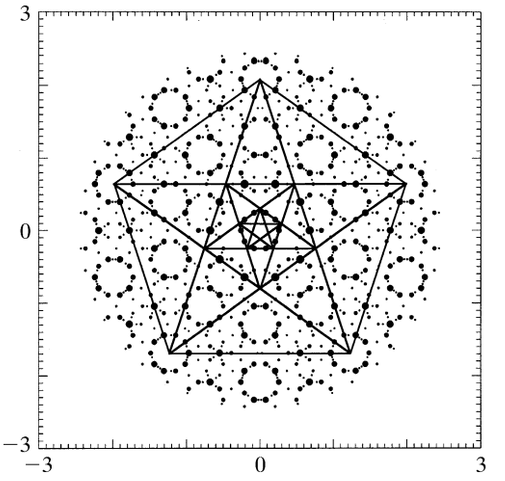


Fig. 4.6.3.25. Pentagrammal relationships between scaling symmetry-related negative structure factors $F(\mathbf{H})$ of the Penrose tiling (edge length $a_r = 4.04$ Å) in parallel space. The magnitudes of the structure factors are indicated by the diameters of the filled circles.

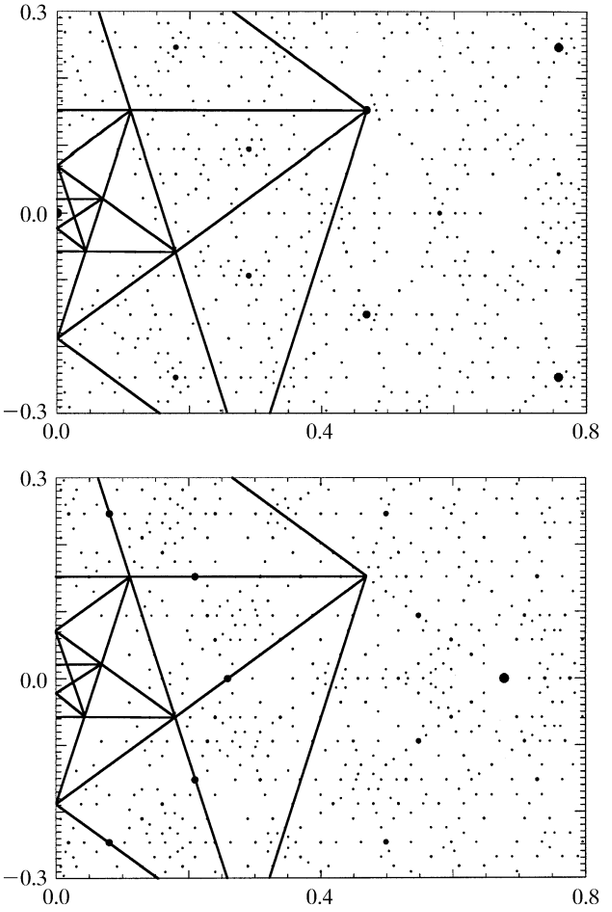


Fig. 4.6.3.26. Pentagrammal relationships between scaling symmetry-related structure factors $F(\mathbf{H})$ of the Penrose tiling (edge length $a_r = 4.04$ Å) in parallel space. Enlarged sections of Figs. 4.6.3.24 (above) and 4.6.3.25 (below) are shown.

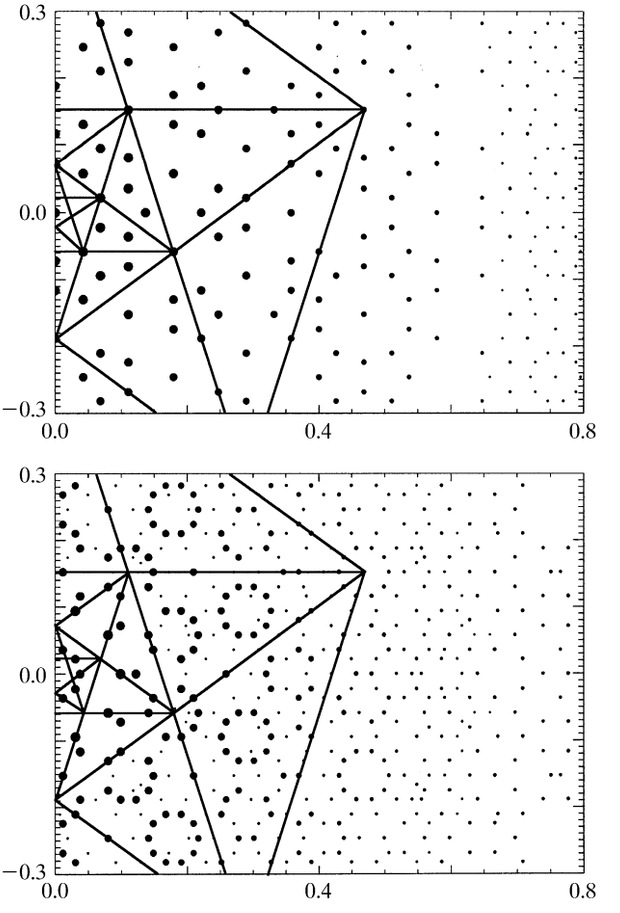


Fig. 4.6.3.27. Pentagrammal relationships between scaling symmetry-related structure factors $F(\mathbf{H})$ of the Penrose tiling (edge length $a_r = 4.04$ Å) in perpendicular space. Enlarged sections of positive (above) and negative structure factors (below) are shown.

$$\begin{pmatrix} \mathbf{a}_1^* \\ \mathbf{a}_2^* \\ \mathbf{a}_3^* \\ \mathbf{a}_4^* \\ \mathbf{a}_5^* \\ \mathbf{a}_6^* \end{pmatrix} = a^* \begin{pmatrix} 0 & 0 & 1 \\ \sin\theta\cos(4\pi/5) & \sin\theta\sin(4\pi/5) & \cos\theta \\ \sin\theta\cos(6\pi/5) & \sin\theta\sin(6\pi/5) & \cos\theta \\ \sin\theta\cos(8\pi/5) & \sin\theta\sin(8\pi/5) & \cos\theta \\ \sin\theta & 0 & \cos\theta \\ \sin\theta\cos(2\pi/5) & \sin\theta\sin(2\pi/5) & \cos\theta \end{pmatrix} \begin{pmatrix} \mathbf{e}_1^V \\ \mathbf{e}_2^V \\ \mathbf{e}_3^V \end{pmatrix},$$

where $\mathbf{e}_i^V$ are Cartesian basis vectors. Thus, from the number of independent reciprocal-basis vectors needed to index the Bragg reflections with integer numbers, the dimension of the embedding space has to be six. The vector components refer to a Cartesian coordinate system (V basis) in the physical (parallel) space.

The set $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{6} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ of all diffraction vectors remains invariant under the action of the symmetry operators of the icosahedral point group $K = m\bar{3}\bar{5}$. The symmetry-adapted matrix representations for the point-group generators, one five-

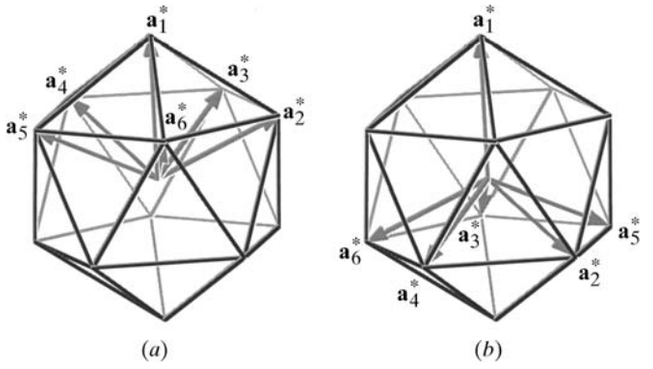


Fig. 4.6.3.28. Perspective (*a*) parallel- and (*b*) perpendicular-space views of the reciprocal basis of the 3D Penrose tiling. The six rationally independent vectors $\mathbf{a}_i^*$ point to the edges of an icosahedron.

fold rotation α, a threefold rotation β and the inversion operation γ, can be written in the form

$$\Gamma(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_D, \Gamma(\beta) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \bar{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}_D,$$

$$\Gamma(\gamma) = \begin{pmatrix} \bar{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \bar{1} \end{pmatrix}_D.$$

Block-diagonalization of these reducible symmetry matrices decomposes them into non-equivalent irreducible representations. These can be assigned to the two orthogonal subspaces forming the 6D embedding space $\mathbf{V} = \mathbf{V}^{\parallel} \oplus \mathbf{V}^{\perp}$, the 3D parallel (physical) subspace $\mathbf{V}^{\parallel}$ and the perpendicular 3D subspace $\mathbf{V}^{\perp}$. Thus, using $W\Gamma W^{-1} = \Gamma^{\rm red} = \Gamma^{\parallel} \oplus \Gamma^{\perp}$, we obtain

$$\Gamma(\alpha) = \left(\begin{array}{ccc|ccc} \cos(2\pi/5) & -\sin(2\pi/5) & 0 & 0 & 0 & 0 \\ \sin(2\pi/5) & \cos(2\pi/5) & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \cos(4\pi/5) & -\sin(4\pi/5) & 0 \\ 0 & 0 & 0 & \sin(4\pi/5) & \cos(4\pi/5) & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right)_V = \left(\begin{array}{c|c} \Gamma^{\parallel} & 0 \\ \hline 0 & \Gamma^{\perp} \end{array}\right)_V,$$

where

$$W = a^* \begin{pmatrix} 0 & sc4 & sc6 & sc8 & s & sc2 \\ 0 & ss4 & ss6 & ss8 & 0 & ss2 \\ 1 & c & c & c & c & c \\ \hline 0 & -sc8 & -sc2 & -sc6 & -s & -sc4 \\ 0 & -ss8 & -ss2 & -ss6 & 0 & -ss4 \\ 1 & -c & -c & -c & -c & -c \end{pmatrix}_V,$$

$c = \cos\theta$, $s = \sin\theta$, $scn = \sin\theta\cos(n\pi/5)$, $ssn = \sin\theta\sin(n\pi/5)$. The column vectors of the matrix W give the parallel- (above the partition line) and perpendicular-space components (below the partition line) of a reciprocal basis in $\mathbf{V}$. Thus, W can be rewritten using the physical-space reciprocal basis defined above and an arbitrary constant c,

$$W = \begin{pmatrix} \mathbf{a}_1^* & \mathbf{a}_2^* & \mathbf{a}_3^* & \mathbf{a}_4^* & \mathbf{a}_5^* & \mathbf{a}_6^* \\ c\mathbf{a}_1^* & -c\mathbf{a}_4^* & -c\mathbf{a}_6^* & -c\mathbf{a}_3^* & -c\mathbf{a}_5^* & -c\mathbf{a}_2^* \end{pmatrix} = \begin{pmatrix} \mathbf{d}_1^* & \mathbf{d}_2^* & \mathbf{d}_3^* & \mathbf{d}_4^* & \mathbf{d}_5^* & \mathbf{d}_6^* \end{pmatrix},$$

yielding the reciprocal basis $\mathbf{d}_i^*, i = 1, \ldots, 6$, in the 6D embedding space (D space)

$$\mathbf{d}_1^* = a^* \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ c \end{pmatrix} \text{ and } \mathbf{d}_i^* = a^* \begin{pmatrix} \sin\theta\cos(2\pi i/5) \\ \sin\theta\sin(2\pi i/5) \\ \cos\theta \\ -c\sin\theta\cos(4\pi i/5) \\ -c\sin\theta\sin(4\pi i/5) \\ -c\cos\theta \end{pmatrix}, i = 2, \ldots, 6.$$

The 6×6 symmetry matrices can each be decomposed into two 3×3 matrices. The first one, $\Gamma^{\parallel}$, acts on the parallel-space component, the second one, $\Gamma^{\perp}$, on the perpendicular-space component. In the case of $\Gamma(\alpha)$, the coupling factor between a rotation in parallel and perpendicular space is 2. Thus a $2\pi/5$ rotation in physical space is related to a $4\pi/5$ rotation in perpendicular space (Figs. 4.6.3.28 and 4.6.3.29).

With the condition $\mathbf{d}_i \cdot \mathbf{d}_j^* = \delta_{ij}$, the basis in direct 6D space is obtained:

$$\mathbf{d}_1 = \frac{1}{2a^*} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1/c \end{pmatrix} \text{ and } \mathbf{d}_i = \frac{1}{2a^*} \begin{pmatrix} \sin\theta\cos(2\pi i/5) \\ \sin\theta\sin(2\pi i/5) \\ \cos\theta \\ -(1/c)\sin\theta\cos(4\pi i/5) \\ -(1/c)\sin\theta\sin(4\pi i/5) \\ -(1/c)\cos\theta \end{pmatrix}, \quad i = 2, \ldots, 6.$$

The metric tensors G, G^* are of the type

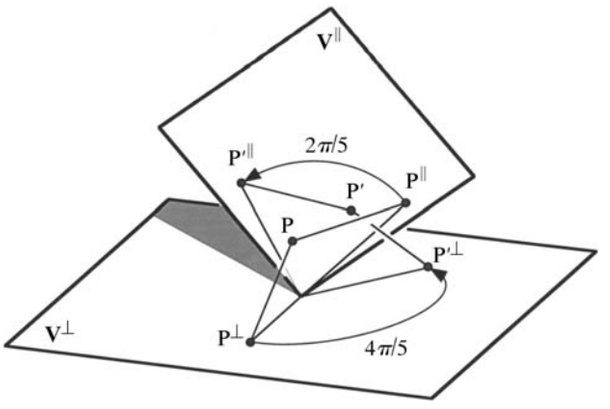


Fig. 4.6.3.29. Schematic representation of a rotation in 6D space. The point P is rotated to P′. The component rotations in parallel and perpendicular space are illustrated.

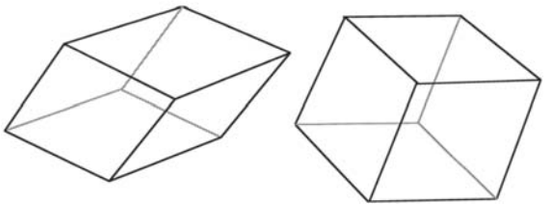


Fig. 4.6.3.30. The two unit tiles of the 3D Penrose tiling: a prolate [$\alpha_p = \arccos(5^{-1/2}) \simeq 63.44^\circ$] and an oblate ($\alpha_o = 180^\circ - \alpha_p$) rhombohedron with equal edge lengths a_r.

$$\begin{pmatrix} A & B & B & B & B & B \\ B & A & B & -B & -B & B \\ B & B & A & B & -B & -B \\ B & -B & B & A & B & -B \\ B & -B & -B & B & A & B \\ B & B & -B & -B & B & A \end{pmatrix},$$

with $A = (1 + c^2)a^{*2}$, $B = [(5)^{1/2}/5](1 - c^2)a^{*2}$ for the reciprocal space and $A = (1 + c^2)/[4(ca^*)^2]$, $B = [(5)^{1/2}(c^2 - 1)]/[20(ca^*)^2]$ for the direct space. For $c = 1$ we obtain hypercubic direct and reciprocal 6D lattices.

The lattice parameters in reciprocal and direct space are $d_i^* = a^*(2)^{1/2}$ and $d_i = 1/[(2)^{1/2}a^*]$ with $i = 1, \ldots, 6$, respectively. The volume of the 6D unit cell can be calculated from the metric tensor G. For $c = 1$ it is simply $V = [\det(G)]^{1/2} = \{1/[(2)^{1/2}a^*]\}^6$.

The best known example of a 3D quasiperiodic structure is the canonical 3D *Penrose tiling* (see Janssen, 1986). It can be constructed from two unit tiles: a prolate and an oblate rhombohedron with equal edge lengths a_r (Fig. 4.6.3.30). Each face of the rhombohedra is a rhomb with acute angles $\alpha_r = \arccos[1/(5)^{1/2}] \simeq 63.44^\circ$. Their volumes are $V_p = (4/5)a_r^3 \sin(2\pi/5)$, $V_o = (4/5)a_r^3 \sin(\pi/5) = V_p/\tau$, and their frequencies $\nu_p{:}\nu_o = \tau{:}1$. The resulting point density (number of vertices per unit volume) is $\rho_p = (\tau + 1)/(\tau V_p + V_o) = (\tau/a_r^3)\sin(2\pi/5)$. Ten prolate and ten oblate rhombohedra can be packed to form a rhombic triacontahedron. The icosahedral symmetry of this zonohedron is broken by the many possible decompositions into the rhombohedra. Removing one zone of the triacontahedron gives a rhomb-icosahedron consisting of five prolate and five oblate rhombohedra. Again, the singular fivefold axis of the rhomb-icosahedron is broken by the decomposition into rhombohedra. Removing one zone again gives a rhombic dodecahedron consisting of two prolate and two oblate rhombohedra. Removing the last remaining zone leads finally to a single prolate rhombohedron. Using these zonohedra as elementary clusters, a matching rule can be derived for the 3D construction of the 3D Penrose tiling (Levine & Steinhardt, 1986; Socolar & Steinhardt, 1986).

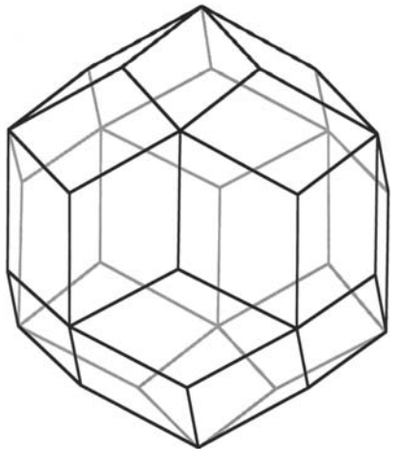


Fig. 4.6.3.31. Atomic surface of the 3D Penrose tiling in the 6D hypercubic description. The projection of the 6D hypercubic unit cell upon $\mathbf{V}^\perp$ gives a rhombic triacontahedron.

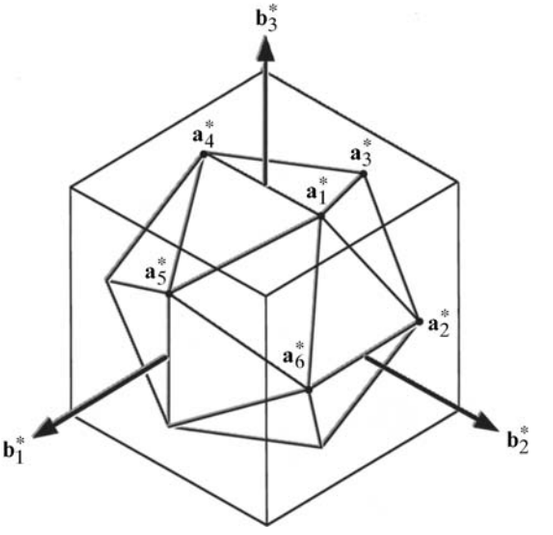


Fig. 4.6.3.32. Perspective parallel-space view of the two alternative reciprocal bases of the 3D Penrose tiling: the cubic and the icosahedral setting, represented by the bases $\mathbf{b}_i^*$, $i = 1, \ldots, 3$, and $\mathbf{a}_i^*$, $i = 1, \ldots, 6$, respectively.

The 3D Penrose tiling can be embedded in the 6D space as shown above. The 6D hypercubic lattice is decorated on the lattice nodes with 3D triacontahedra obtained from the projection of a 6D unit cell upon the perpendicular space $\mathbf{V}^\perp$ (Fig. 4.6.3.31). Thus the edge length of the rhombs covering the triacontahedron is equivalent to the length $\pi^\perp(\mathbf{d}_i) = 1/2a^*$ of the perpendicular-space component of the vectors spanning the 6D hypercubic lattice $\Sigma = \{\mathbf{r} = \sum_{i=1}^6 n_i\mathbf{d}_i | n_i \in \mathbb{Z}\}$.

4.6.3.3.3.1. Indexing

There are several indexing schemes in use. The generic one uses a set of six rationally independent reciprocal-basis vectors pointing to the corners of an icosahedron, $\mathbf{a}_1^* = a^*(0, 0, 1)$, $\mathbf{a}_i^* = a^*[\sin\theta\cos(2\pi i/5), \sin\theta\sin(2\pi i/5), \cos\theta]$, $i = 2, \ldots, 6$, $\sin\theta = 2/(5)^{1/2}$, $\cos\theta = 1/(5)^{1/2}$, with $\theta \simeq 63.44^\circ$, the angle between two neighbouring fivefold axes (setting 1) (Fig. 4.6.3.28). In this case, the physical-space basis corresponds to a simple projection of the 6D reciprocal basis $\mathbf{d}_i^*$, $i = 1, \ldots, 6$. Sometimes, the same set of six reciprocal-basis vectors is referred to a differently oriented Cartesian reference system (C basis, with basis vectors $\mathbf{e}_i$ along the twofold axes) (Bancel *et al.*, 1985). The reciprocal basis is

$$\begin{pmatrix} \mathbf{a}_1^* \\ \mathbf{a}_2^* \\ \mathbf{a}_3^* \\ \mathbf{a}_4^* \\ \mathbf{a}_5^* \\ \mathbf{a}_6^* \end{pmatrix} = \frac{a^*}{(1+\tau^2)^{1/2}} \begin{pmatrix} 0 & 1 & \tau \\ -1 & \tau & 0 \\ -\tau & 0 & 1 \\ 0 & -1 & \tau \\ \tau & 0 & 1 \\ 1 & \tau & 0 \end{pmatrix}_C \begin{pmatrix} \mathbf{e}_1^C \\ \mathbf{e}_2^C \\ \mathbf{e}_3^C \end{pmatrix}.$$

An alternate way of indexing is based on a 3D set of cubic reciprocal-basis vectors $\mathbf{b}_i^*$, $i = 1, \ldots, 3$ (setting 2) (Fig. 4.6.3.32):

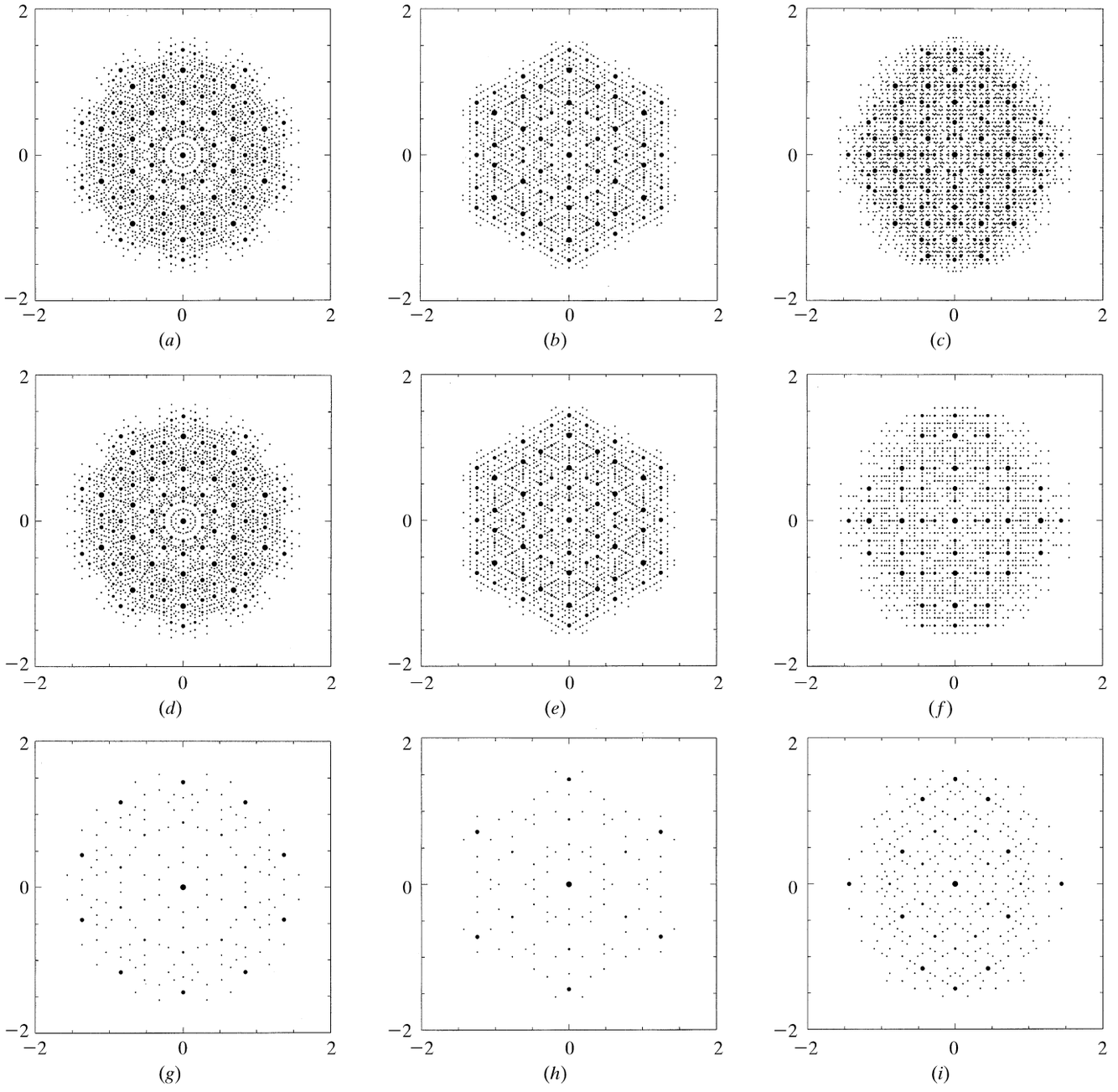


Fig. 4.6.3.33. Physical-space diffraction patterns of the 3D Penrose tiling decorated with point atoms (edge lengths of the Penrose unit rhombohedra $a_r = 5.0$ Å). Sections with five-, three- and twofold symmetry are shown for the primitive 6D analogue of Bravais type P in (a, b, c), the body-centred 6D analogue to Bravais type I in (d, e, f) and the face-centred 6D analogue to Bravais type F in (g, h, i). All reflections are shown within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $-6 \le h_i \le 6, i = 1, \ldots, 6$.

$$\begin{pmatrix} \mathbf{b}_1^* \\ \mathbf{b}_2^* \\ \mathbf{b}_3^* \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} 0 & \bar{1} & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & \bar{1} & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}_D \begin{pmatrix} \mathbf{a}_1^* \\ \mathbf{a}_2^* \\ \mathbf{a}_3^* \\ \mathbf{a}_4^* \\ \mathbf{a}_5^* \\ \mathbf{a}_6^* \end{pmatrix}$$

$$= \frac{a^*}{(1+\tau^2)^{1/2}} \begin{pmatrix} \mathbf{e}_1^C \\ \mathbf{e}_2^C \\ \mathbf{e}_3^C \end{pmatrix}.$$

The Cartesian C basis is related to the V basis by a $\theta/2$ rotation around $[100]_C$, yielding $[001]_V$, followed by a $\pi/10$ rotation around $[001]_C$:

$$\begin{pmatrix} \mathbf{e}_1^C \\ \mathbf{e}_2^C \\ \mathbf{e}_3^C \end{pmatrix} = \begin{pmatrix} \cos(\pi/10) & \sin(\pi/10) & 0 \\ -\cos(\theta/2)\sin(\pi/10) & \cos(\theta/2)\cos(\pi/10) & \sin(\theta/2) \\ \sin(\theta/2)\sin(\pi/10) & -\sin(\theta/2)\cos(\pi/10) & \cos(\theta/2) \end{pmatrix}_V \begin{pmatrix} \mathbf{e}_1^V \\ \mathbf{e}_2^V \\ \mathbf{e}_3^V \end{pmatrix}.$$

Thus, indexing the diffraction pattern of an icosahedral phase with integer indices, one obtains for setting 1 $\mathbf{H} = \sum_{i=1}^{6} h_i \mathbf{a}_i^*, h_i \in \mathbb{Z}$. These indices $(h_1\ h_2\ h_3\ h_4\ h_5\ h_6)$ transform into the second setting to $(h/h'\ k/k'\ l/l')$ with the fractional cubic indices $h_1^c = h + h'\tau$, $h_2^c = k + k'\tau$, $h_3^c = l + l'\tau$. The transformation matrix is

Table 4.6.3.2. *3D point groups of order k describing the diffraction symmetry and corresponding 6D decagonal space groups with reflection conditions (see Levitov & Rhyner, 1988; Rokhsar et al., 1988)*

3D point group	k	5D space group	Reflection condition
$\frac{2}{m}\bar{3}\bar{5}$	120	$P\frac{2}{m}\bar{3}\bar{5}$	No condition
		$P\frac{2}{n}\bar{3}\bar{5}$	$h_1h_2\overline{h_1h_2}h_5h_6$: $h_5 - h_6 = 2n$
		$I\frac{2}{m}\bar{3}\bar{5}$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i=1}^{6} h_i = 2n$
		$F\frac{2}{m}\bar{3}\bar{5}$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i\neq j=1}^{6} h_i + h_j = 2n$
		$F\frac{2}{n}\bar{3}\bar{5}$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i\neq j=1}^{6} h_i + h_j = 2n$ $h_1h_2\overline{h_1h_2}h_5h_6$: $h_5 - h_6 = 4n$
235	60	$P235$	No condition
		$P235_1$	$h_1h_2h_2h_2h_2h_2$: $h_1 = 5n$
		$I235$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i=1}^{6} h_i = 2n$
		$I235_1$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i=1}^{6} h_i = 2n$ $h_1h_2h_2h_2h_2h_2$: $h_1 + 5h_2 = 10n$
		$F235$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i\neq j=1}^{6} h_i + h_j = 2n$
		$F235_1$	$h_1h_2h_3h_4h_5h_6$: $\sum_{i\neq j=1}^{6} h_i + h_j = 2n$ $h_1h_2h_2h_2h_2h_2$: $h_1 + 5h_2 = 10n$

$$\begin{pmatrix} h \\ h' \\ k \\ k' \\ l \\ l' \end{pmatrix}_C = \begin{pmatrix} 0 & \bar{1} & 0 & 0 & 0 & 1 \\ 0 & 0 & \bar{1} & 0 & 1 & 0 \\ 1 & 0 & 0 & \bar{1} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{pmatrix}_D = \begin{pmatrix} h_6 - h_2 \\ h_5 - h_3 \\ h_1 - h_4 \\ h_6 + h_2 \\ h_5 + h_3 \\ h_1 + h_4 \end{pmatrix}_D .$$

4.6.3.3.3.2. Diffraction symmetry

The diffraction symmetry of icosahedral phases can be described by the Laue group $K = m\bar{3}\bar{5}$. The set of all vectors $\mathbf{H}$ forms a Fourier module $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{6} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ of rank 6 in physical space. Consequently, it can be considered as a projection from a 6D reciprocal lattice, $M^* = \pi^{\parallel}(\Sigma^*)$. The parallel and perpendicular reciprocal-space sections of the 3D Penrose tiling decorated with equal point scatterers on its vertices are shown in Figs. 4.6.3.33 and 4.6.3.34. The diffraction pattern in perpendicular space is the Fourier transform of the triacontahedron. All Bragg reflections within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ are depicted. Without intensity-truncation limit, the diffraction pattern would be densely filled with discrete Bragg reflections.

The 6D icosahedral space groups that are relevant to the description of icosahedral phases (six symmorphous and five non-symmorphous groups) are listed in Table 4.6.3.2. These space groups are a subset of all 6D icosahedral space groups fulfilling the condition that the 6D point groups they are associated with are isomorphous to the 3D point groups $\frac{2}{m}\bar{3}\bar{5}$ and 235 describing the diffraction symmetry. From 826 6D (analogues to) Bravais groups (Levitov & Rhyner, 1988), only three fulfil the condition that the projection of the 6D hypercubic lattice upon the 3D physical space is compatible with the icosahedral point groups $\frac{2}{m}\bar{3}\bar{5}$, 235: the primitive hypercubic Bravais lattice P, the body-centred Bravais lattice I with translation 1/2(111111), and the face-centred Bravais lattice F with translations 1/2(110000) + 14 further cyclic permutations. Hence, the I lattice is twofold primitive (*i.e.* it contains two vertices per unit cell) and the F lattice is 16-fold primitive. The orientation of the symmetry elements in the 6D space is defined by the isomorphism of the 3D and 6D point groups. The action of the fivefold rotation, however, is different in the subspaces $\mathbf{V}^{\parallel}$ and $\mathbf{V}^{\perp}$: a rotation of $2\pi/5$ in $\mathbf{V}^{\parallel}$ is correlated with a rotation of $4\pi/5$ in $\mathbf{V}^{\perp}$. The reflection and inversion operations are equivalent in both subspaces.

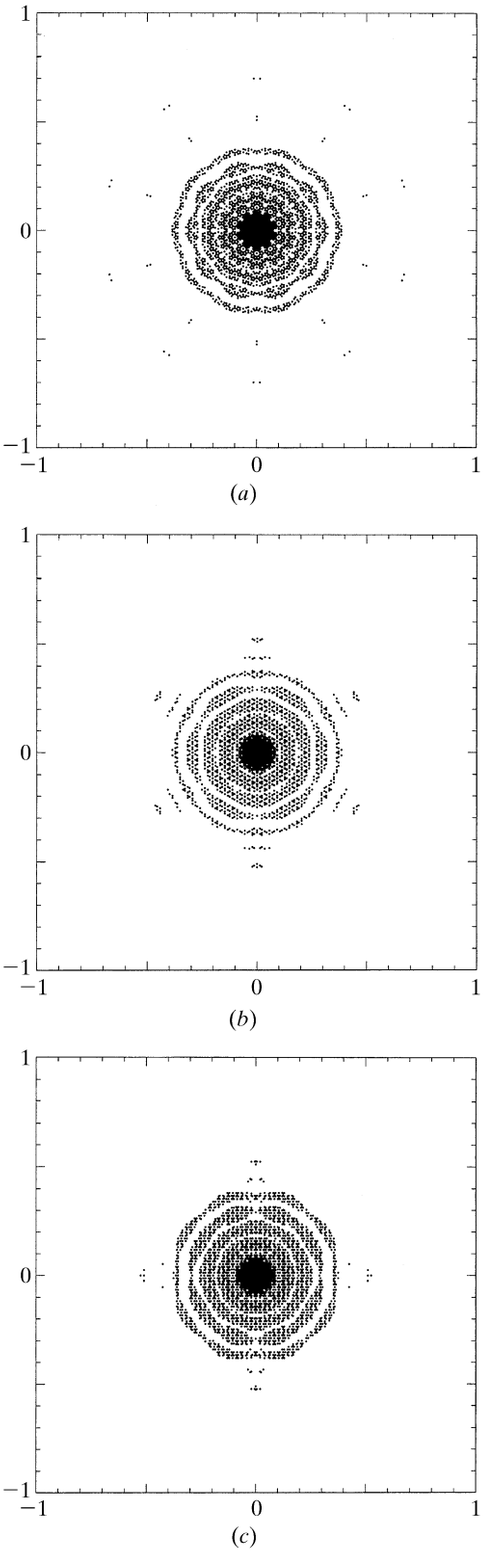


Fig. 4.6.3.34. Perpendicular-space diffraction patterns of the 3D Penrose tiling decorated with point atoms (edge lengths of the Penrose unit rhombohedra a_r = 5.0 Å). Sections with (*a*) five-, (*b*) three- and (*c*) twofold symmetry are shown for the primitive 6D analogue of Bravais type P. All reflections are shown within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $-6 \leq h_i \leq 6, i = 1, \ldots, 6$.

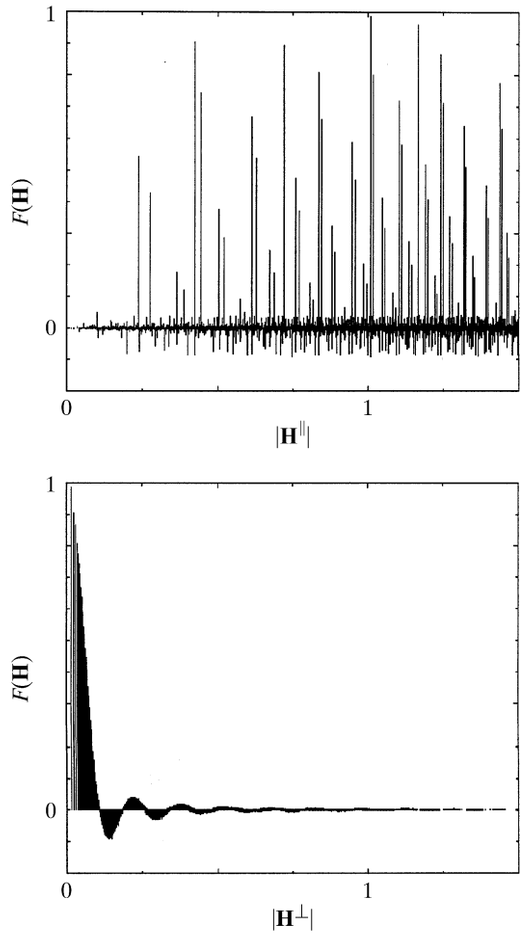


Fig. 4.6.3.35. Radial distribution function of the structure factors $F(\mathbf{H})$ of the 3D Penrose tiling (edge lengths of the Penrose unit rhombohedra $a_r = 5.0$ Å) decorated with point atoms as a function of $|\mathbf{H}^{\parallel}|$ (above) and $|\mathbf{H}^{\perp}|$ (below). All reflections are shown within $10^{-6}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $-6 \le h_i \le 6, i = 1, \ldots, 6$.

4.6.3.3.3.3. Structure factor

The structure factor of the icosahedral phase corresponds to the Fourier transform of the 6D unit cell,

$$F(\mathbf{H}) = \sum_{k=1}^{N} f_k(\mathbf{H}^{\parallel}) T_k(\mathbf{H}^{\parallel}, \mathbf{H}^{\perp}) g_k(\mathbf{H}^{\perp}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_k),$$

with 6D diffraction vectors $\mathbf{H} = \sum_{i=1}^{6} h_i \mathbf{d}_i^*$, parallel-space atomic scattering factor $f_k(H^{\parallel})$, temperature factor $T_k(\mathbf{H}^{\parallel}, \mathbf{H}^{\perp})$ and perpendicular-space geometric form factor $g_k(\mathbf{H}^{\perp})$. $T_k(\mathbf{H}^{\parallel}, \mathbf{0})$ is equivalent to the conventional Debye–Waller factor and $T_k(\mathbf{0}, \mathbf{H}^{\perp})$ describes random fluctuations in perpendicular space. These fluctuations cause characteristic jumps of vertices (*phason flips*) in the physical space. Even random phason flips map the vertices onto positions that can still be described by physical-space vectors of the type $\mathbf{r} = \sum_{i=1}^{6} n_i \mathbf{a}_i$. Consequently, the set $M = \{\mathbf{r} = \sum_{i=1}^{6} n_i \mathbf{a}_i | n_i \in \mathbb{Z}\}$ of all possible vectors forms a $\mathbb{Z}$ module. The shape of the atomic surfaces corresponds to a selection rule for the positions actually occupied. The geometric form factor $g_k(\mathbf{H}^{\perp})$ is equivalent to the Fourier transform of the *atomic surface*, *i.e.* the 3D perpendicular-space component of the 6D *hyperatoms*.

For the example of the canonical 3D Penrose tiling, $g_k(\mathbf{H}^{\perp})$ corresponds to the Fourier transform of a triacontahedron:

$$g_k(\mathbf{H}^{\perp}) = (1/A_{\mathrm{UC}}^{\perp}) \int_{A_k} \exp(2\pi i \mathbf{H}^{\perp} \cdot \mathbf{r})\, \mathrm{d}\mathbf{r},$$

where $A_{\mathrm{UC}}^{\perp}$ is the volume of the 6D unit cell projected upon $\mathbf{V}^{\perp}$ and A_k is the volume of the triacontahedron. $A_{\mathrm{UC}}^{\perp}$ and A_k are equal in the present case and amount to the volumes of ten prolate and ten oblate rhombohedra: $A_{\mathrm{UC}}^{\perp} = 8a_r^3[\sin(2\pi/5) + \sin(\pi/5)]$. Evaluating the integral by decomposing the triacontahedron into trigonal pyramids, each one directed from the centre of the triacontahedron to three of its corners given by the vectors $\mathbf{e}_i$, $i = 1, \ldots, 3$, one obtains

$$g(\mathbf{H}^{\perp}) = (1/A_{\mathrm{UC}}^{\perp}) \sum_{R} g_k(R^T \mathbf{H}^{\perp}),$$

with $k = 1, \ldots, 60$ running over all site-symmetry operations R of the icosahedral group,

$$\begin{aligned} g_k(\mathbf{H}^{\perp}) = -iV_r[&A_2A_3A_4 \exp(iA_1) + A_1A_3A_5 \exp(iA_2) \\ &+ A_1A_2A_6 \exp(iA_3) + A_4A_5A_6] \\ &\times (A_1A_2A_3A_4A_5A_6)^{-1}, \end{aligned}$$

$A_j = 2\pi \mathbf{H}^{\perp} \cdot \mathbf{e}_j$, $j = 1, \ldots, 3$, $A_4 = A_2 - A_3$, $A_5 = A_3 - A_1$, $A_6 = A_1 - A_2$ and $V_r = \mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3)$ the volume of the parallelepiped defined by the vectors $\mathbf{e}_i$, $i = 1, \ldots, 3$ (Yamamoto, 1992*b*).

4.6.3.3.3.4. Intensity statistics

The radial structure-factor distributions of the 3D Penrose tiling decorated with point scatterers are plotted in Fig. 4.6.3.35 as a function of parallel and perpendicular space. The distribution of $|F(\mathbf{H})|$ as a function of their frequencies clearly resembles a centric distribution, as can be expected from the centrosymmetric unit cell. The shape of the distribution function depends on the radius of the limiting sphere in reciprocal space. The number of weak reflections increases as the power 6, that of strong reflections only as the power 3 (strong reflections always have small $\mathbf{H}^{\perp}$ components).

The weighted reciprocal space of the 3D Penrose tiling contains an infinite number of Bragg reflections within a limited region of the physical space. Contrary to the diffraction pattern of a periodic structure consisting of point atoms on the lattice nodes, the Bragg reflections show intensities depending on the perpendicular-space components of their diffraction vectors.

4.6.3.3.3.5. Relationships between structure factors at symmetry-related points of the Fourier image

The weighted 3D reciprocal space $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{6} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ exhibits the icosahedral point symmetry $K = m\bar{3}5$. It is invariant under the action of the scaling matrix S^3:

$$S = \tfrac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \end{pmatrix}_D, \quad S^3 = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & -1 & -1 & 1 \\ 1 & 1 & 2 & 1 & -1 & -1 \\ 1 & -1 & 1 & 2 & 1 & -1 \\ 1 & -1 & -1 & 1 & 2 & 1 \\ 1 & 1 & -1 & -1 & 1 & 2 \end{pmatrix}_D,$$

$$\begin{pmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & -1 & -1 & 1 \\ 1 & 1 & 2 & 1 & -1 & -1 \\ 1 & -1 & 1 & 2 & 1 & -1 \\ 1 & -1 & -1 & 1 & 2 & 1 \\ 1 & 1 & -1 & -1 & 1 & 2 \end{pmatrix}_D \begin{pmatrix} \mathbf{a}_1^* \\ \mathbf{a}_2^* \\ \mathbf{a}_3^* \\ \mathbf{a}_4^* \\ \mathbf{a}_5^* \\ \mathbf{a}_6^* \end{pmatrix} = \tau^3 \begin{pmatrix} \mathbf{a}_1^* \\ \mathbf{a}_2^* \\ \mathbf{a}_3^* \\ \mathbf{a}_4^* \\ \mathbf{a}_5^* \\ \mathbf{a}_6^* \end{pmatrix}.$$

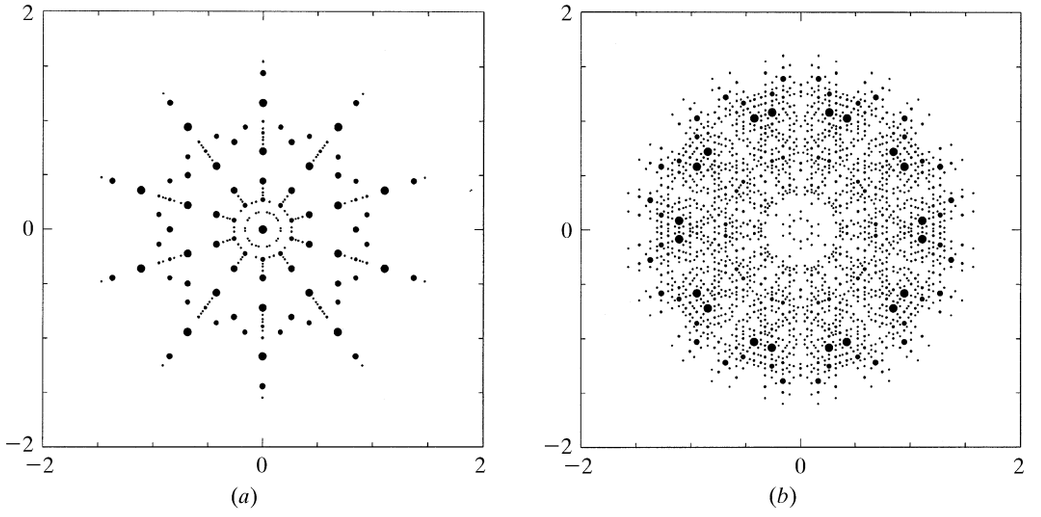


Fig. 4.6.3.36. Parallel-space distribution of (*a*) positive and (*b*) negative structure factors of the 3D Penrose tiling of the 6D *P* lattice type decorated with point atoms (edge lengths of the Penrose unit rhombohedra a_r = 5.0 Å). The magnitudes of the structure factors are indicated by the diameters of the filled circles. All reflections are shown within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $-6 \le h_i \le 6, i = 1, \ldots, 6$.

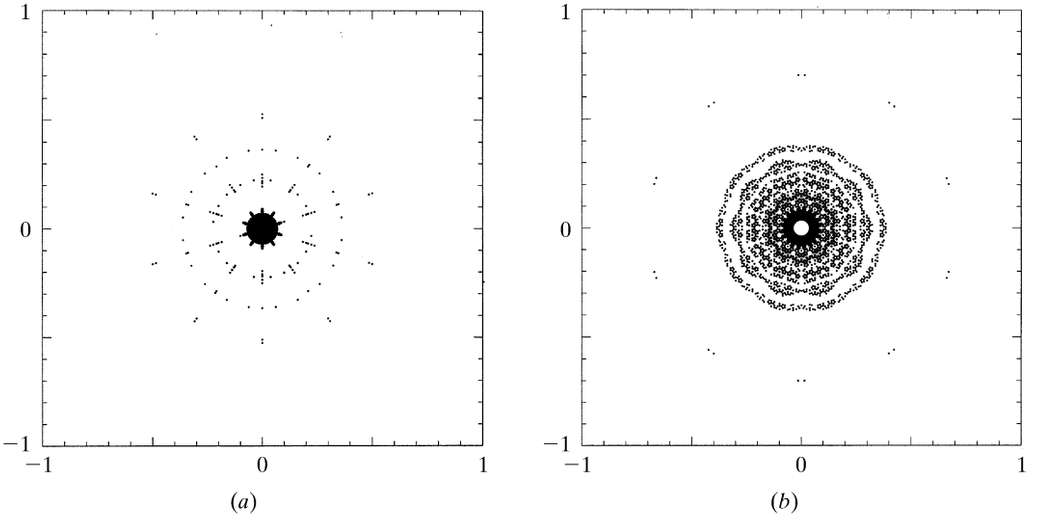


Fig. 4.6.3.37. Perpendicular-space distribution of (*a*) positive and (*b*) negative structure factors of the 3D Penrose tiling of the 6D *P* lattice type decorated with point atoms (edge lengths of the Penrose unit rhombohedra a_r = 5.0 Å). The magnitudes of the structure factors are indicated by the diameters of the filled circles. All reflections are shown within $10^{-4}|F(\mathbf{0})|^2 < |F(\mathbf{H})|^2 < |F(\mathbf{0})|^2$ and $-6 \le h_i \le 6, i = 1, \ldots, 6$.

Table 4.6.4.1. *Intensity statistics of the Fibonacci chain for a total of 161 322 reflections with* $-1000 \le h_i \le 1000$ *and* $0 \le \sin\theta/\lambda \le 2$ Å^{-1}

In the upper line, the number of reflections in the respective interval is given; in the lower line the partial sums $\sum I(\mathbf{H})$ of the intensities $I(\mathbf{H})$ are given as a percentage of the total diffracted intensity. The $F(00)$ reflection is not included in the sums.

	$F(\mathbf{H})/F(\mathbf{H})_{max} \ge 0.1$	$0.1 > F(\mathbf{H})/F(\mathbf{H})_{max} \ge 0.01$	$0.01 > F(\mathbf{H})/F(\mathbf{H})_{max} \ge 0.001$	$F(\mathbf{H})/F(\mathbf{H})_{max} < 0.001$
$0 \le \sin\theta/\lambda \le 0.2$ Å^{-1}	17	148	1505	14 511
$\sum I(\mathbf{H})$	52.53%	2.56%	0.27%	0.03%
$0.2 \le \sin\theta/\lambda \le 0.4$ Å^{-1}	11	107	1066	14 998
$\sum I(\mathbf{H})$	27.03%	2.03%	0.19%	0.02%
$0.4 \le \sin\theta/\lambda \le 0.6$ Å^{-1}	9	64	654	15 456
$\sum I(\mathbf{H})$	9.84%	0.96%	0.12%	0.01%
$0.6 \le \sin\theta/\lambda \le 0.8$ Å^{-1}	6	27	326	15 823
$\sum I(\mathbf{H})$	2.94%	0.34%	0.07%	0.01%
$0.8 \le \sin\theta/\lambda \le 2$ Å^{-1}	1	35	338	96 720
$\sum I(\mathbf{H})$	0.23%	0.79%	0.06%	0.01%
Total sum	44	381	3389	157 508
	92.57%	6.67%	0.70%	0.06%

The scaling transformation $(S^{-3})^T$ leaves a primitive 6D reciprocal lattice invariant as can easily be seen from its application on the indices:

$$\begin{pmatrix} h_1' \\ h_2' \\ h_3' \\ h_4' \\ h_5' \\ h_6' \end{pmatrix} = \begin{pmatrix} -2 & 1 & 1 & 1 & 1 & 1 \\ 1 & -2 & 1 & -1 & -1 & 1 \\ 1 & 1 & -2 & 1 & -1 & -1 \\ 1 & -1 & 1 & -2 & 1 & -1 \\ 1 & -1 & -1 & 1 & -2 & 1 \\ 1 & 1 & -1 & -1 & 1 & -2 \end{pmatrix}_D \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{pmatrix}.$$

The matrix $(S^{-1})^T$ leaves $M^* = \{\mathbf{H}^{\parallel} = \sum_{i=1}^{6} h_i \mathbf{a}_i^* | h_i \in \mathbb{Z}\}$ invariant,

$$\begin{pmatrix} h_1' \\ h_2' \\ h_3' \\ h_4' \\ h_5' \\ h_6' \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 \end{pmatrix}_D \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{pmatrix},$$

for any $\mathbf{H} = \sum_{i=1}^{6} h_i \mathbf{d}_i^*$ with h_i all even or all odd, corresponding to a 6D face-centred hypercubic lattice. In a second case the sum $\sum_{i=1}^{6} h_i$ is even, corresponding to a 6D body-centred hypercubic lattice. Block-diagonalization of the matrix S decomposes it into two irreducible representations. With $WSW^{-1} = S_V = S_V^{\parallel} \oplus S_V^{\perp}$ we obtain

$$S_V = \left(\begin{array}{ccc|ccc} \tau & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau & 0 & 0 & 0 & 0 \\ 0 & 0 & \tau & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1/\tau & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/\tau & 0 \\ 0 & 0 & 0 & 0 & 0 & -1/\tau \end{array}\right)_V = \left(\begin{array}{c|c} S^{\parallel} & 0 \\ \hline 0 & S^{\perp} \end{array}\right)_V,$$

the scaling properties in the two 3D subspaces: scaling by a factor τ in parallel space corresponds to a scaling by a factor $(-\tau)^{-1}$ in perpendicular space. For the intensities of the scaled reflections analogous relationships are valid, as discussed for decagonal phases (Figs. 4.6.3.36 and 4.6.3.37, Section 4.6.3.3.2.5).

4.6.4. Experimental aspects of the reciprocal-space analysis of aperiodic crystals

4.6.4.1. Data-collection strategies

Theoretically, aperiodic crystals show an infinite number of reflections within a given diffraction angle, contrary to periodic crystals. The number of reflections to be included in a structure analysis of a *periodic* crystal may be very high (one million for virus crystals, for instance) but there is no ambiguity in the selection of reflections to be collected: all Bragg reflections within a limiting sphere in reciprocal space, usually given by $0 \leq \sin\theta/\lambda \leq 0.7$ Å^{-1}, are used. All reflections, observed and unobserved, are included to fit a reliable structure model.

However, for *aperiodic* crystals it is not possible to collect the infinite number of dense Bragg reflections within $0 \leq \sin\theta/\lambda \leq 0.7$ Å^{-1}. The number of observable reflections within this limiting sphere depends only on the spatial and intensity resolution.

What happens if not all reflections are included in a structure analysis? How important is the contribution of reflections with large perpendicular-space components of the diffraction vector which are weak but densely distributed? These problems are illustrated using the example of the Fibonacci sequence. An infinite model structure consisting of Al atoms with isotropic thermal parameter $B = 1$ Å^2, and distances S = 2.5 Å and L = τS, was used for the calculations (Table 4.6.4.1).

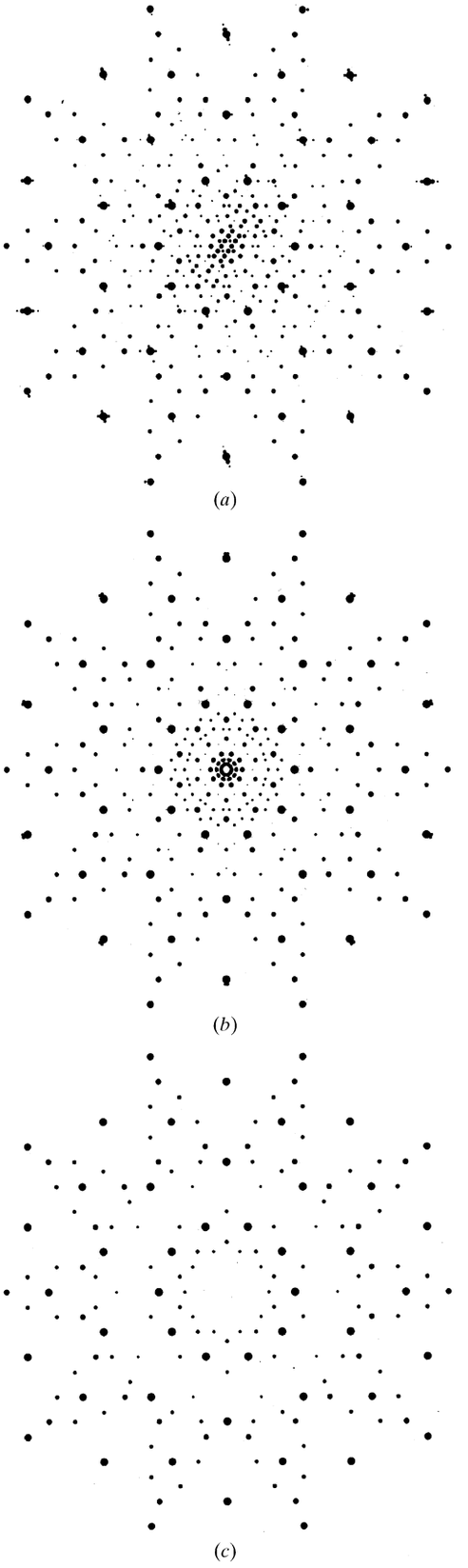


Fig. 4.6.4.1. Simulated diffraction patterns of (*a*) the $\simeq 52$ Å single-crystal approximant of decagonal Al–Co–Ni, (*b*) the fivefold twinned approximant, and (*c*) the decagonal phase itself (Estermann *et al.*, 1994).

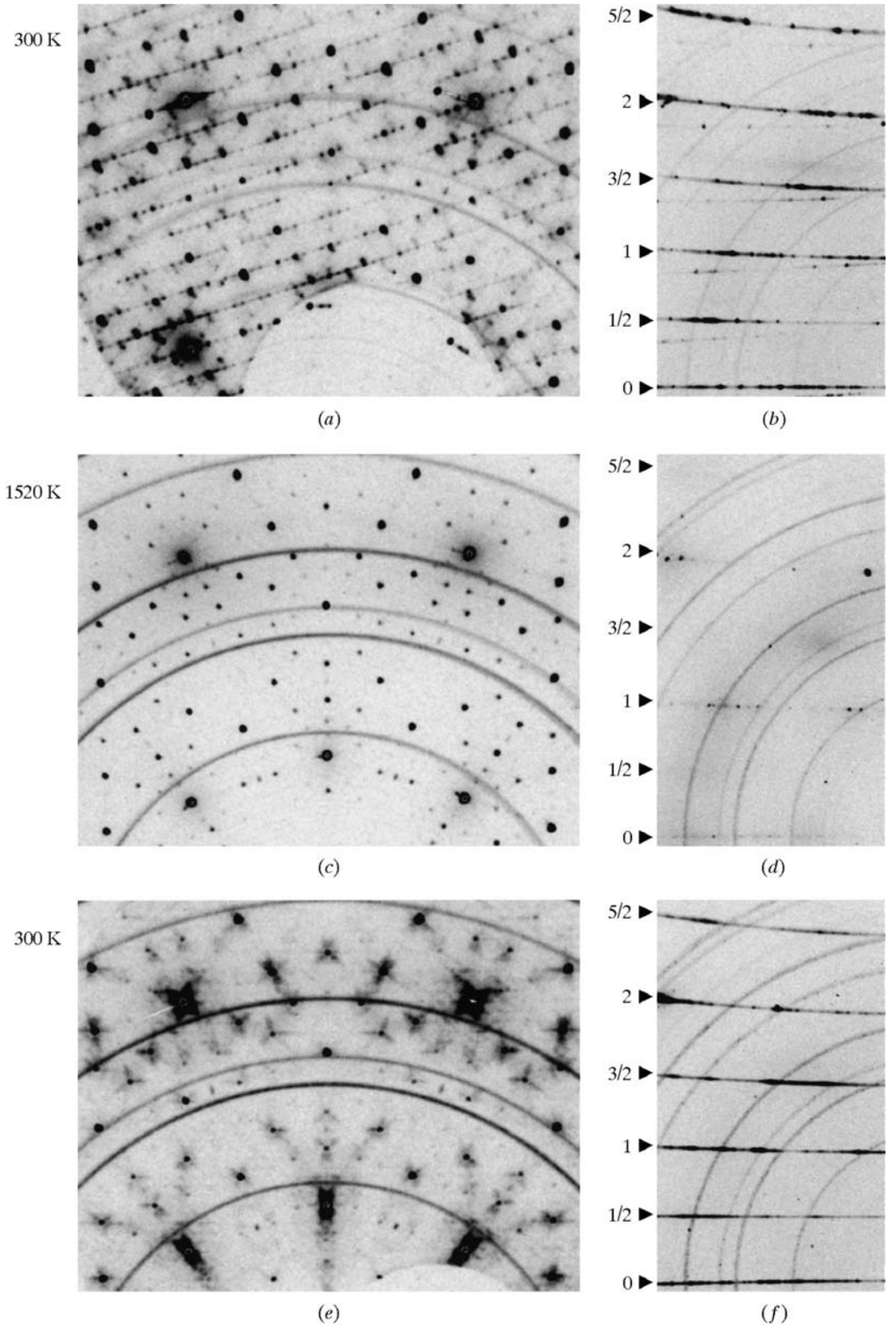


Fig. 4.6.4.2. Zero-layer X-ray diffraction patterns of decaprismatic $Al_{73.5}C_{21.7}Ni_{4.8}$ crystals taken parallel and perpendicular to the crystal axis on an image-plate scanner (Mar Research) at different temperatures. In (*a*) and (*b*), room-temperature (RT) diffraction patterns from a sample quenched after annealing at 1073 K are shown. Reflections from both a crystalline approximant and a decagonal phase are visible. The period along the unique direction in the decagonal phase and the corresponding period in the approximant phase is $\simeq 8$ Å (*b*). At 1520 K, a single-phase decagonal quasicrystal is present with $\simeq 4$ Å fundamental structure (*c*, *d*). In (*e*, *f*), the RT diffraction patterns of the slowly cooled sample indicate a single-phase nanodomain structure with $\simeq 8$ Å periodicity along the unique direction.

It turns out that 92.6% of the total diffracted intensity of 161 322 reflections is included in the 44 strongest reflections and 99.2% in the strongest 425 reflections. It is remarkable, however, that in all the experimental data for icosahedral and decagonal quasicrystals collected so far, rarely more than 20 to 50 reflections along reciprocal-lattice lines corresponding to net planes with Fibonacci-sequence-like distances could be observed. The consequences for structure determinations with such truncated data sets are primarily a lower resolution in perpendicular space than in physical space. This corresponds to a smearing of the hyperatoms in the perpendicular space. For the derivation of the local structure-building elements (clusters) of aperiodic crystals this is only a minor problem: the smeared hyperatoms give rise to split atoms and a biased electron-density distribution. The information on the global aperiodic structure, however, which is contained in the detailed shape of the atomic surfaces, is severely

reduced when using a low-resolution diffraction data set. A combination of high-resolution electron microscopy, lattice imaging and diffraction techniques allows a good characterization of the local and global order even in these cases. For a more detailed analysis of these problems see Steurer (1995).

4.6.4.2. Commensurability versus incommensurability

The question whether an aperiodic crystal is really aperiodic or rather a high-order approximant is of different importance depending on the point of view. As far as real finite crystals are considered, definitions of *periodic* and *aperiodic real crystals* and of *periodic* and *aperiodic perfect crystals* have to be given first. *Real crystals*, despite periodicity or aperiodicity, are the actual samples under investigation. Partial information about their actual structure can be obtained today by imaging methods (scanning tunnelling microscopy, atomic force microscopy, high-resolution transmission electron microscopy, ...). Basically, the real crystal structure can be determined using full diffraction information from Bragg and diffuse scattering. In practice, however, only 'Bragg reflections' are included in a structure analysis. 'Bragg reflections' result from the integration of diffraction intensities from extended volumes around a limited number of Bragg-reflection positions ($\mathbb{Z}$ module). This process of intensity condensation at Bragg points corresponds in direct space to an averaging process. The real crystal structure is projected upon one unit cell in direct space defined by the $\mathbb{Z}$ module in reciprocal space. Generally, the identification of appropriate reciprocal-space metrics is not a problem in the case of crystals. It can be problematic, however, in the case of aperiodic crystals, in particular quasicrystals (see Lancon *et al.*, 1994). The metrics, and to some extent the global order in the case of quasicrystals, are fixed by assigning the reciprocal basis. The spatial resolution of a diffraction experiment defines the accuracy of the resulting metrics. The decision whether the rational number obtained for the relative length of a satellite vector indicates a commensurate or an incommensurate modulation can only be made considering temperature- and pressure-dependent chemical and physical properties of the material. The same is valid for other types of aperiodic crystals.

4.6.4.3. Twinning and nanodomain structures

High-order approximants of quasicrystals often occur in orientationally twinned form or, on a smaller scale, as oriented nanodomain structures. These structures can be identified by electron microscopy, and, in certain cases, also by high-resolution X-ray diffractometry (Kalning *et al.*, 1994). If the intensity and spatial resolution is sufficient, characteristic reflection splitting and diffuse diffraction phenomena can be observed. It has been demonstrated that for the determination of the local structure (structure-building elements) it does not matter greatly whether one uses a data set for a real quasicrystal or one for a twinned approximant (Estermann *et al.*, 1994). Examples of reciprocal-space images of an approximant, a twinned approximant and the related decagonal phase are shown schematically in Fig. 4.6.4.1 and for real samples in Fig. 4.6.4.2.

References

Axel, F. & Gratias, D. (1995). Editors. *Beyond Quasicrystals.* Les Ulis: Les Editions de Physique and Berlin: Springer-Verlag.

Bancel, P. A., Heiney, P. A., Stephens, P. W., Goldman, A. I. & Horn, P. M. (1985). *Structure of rapidly quenched Al–Mn. Phys. Rev. Lett.* **54**, 2422–2425.

Böhm, H. (1977). *Eine erweiterte Theorie der Satellitenreflexe und die Bestimmung der modulierten Struktur des Natriumnitrits.* Habilitation thesis, University of Munster.

Cummins, H. Z. (1990). *Experimental studies of structurally incommensurate crystal phases. Phys. Rep.* **185**, 211–409.

Dehlinger, U. (1927). *Über die Verbreiterung der Debyelinien bei kaltbearbeiteten Metallen. Z. Kristallogr.* **65**, 615–631.

Dräger, J. & Mermin, N. D. (1996). *Superspace groups without the embedding: the link between superspace and Fourier-space crystallography. Phys. Rev. Lett.* **76**, 1489–1492.

Estermann, M., Haibach, T. & Steurer, W. (1994). *Quasicrystal versus twinned approximant: a quantitative analysis with decagonal $Al_{70}Co_{15}Ni_{15}$. Philos. Mag. Lett.* **70**, 379–384.

Goldman, A. I. & Kelton, K. F. (1993). *Quasicrystals and crystalline approximants. Rev. Mod. Phys.* **65**, 213–230.

Gouyet, J. F. (1996). *Physics and Fractal Structures.* Paris: Masson and Berlin: Springer-Verlag.

Hausdorff, F. (1919). *Dimension und äusseres Mass. Math. Ann.* **79**, 157–179.

Hermann, C. (1949). *Kristallographie in Räumen beliebiger Dimensionszahl. I. Die Symmetrieoperationen. Acta Cryst.* **2**, 139–145.

Ishihara, K. N. & Yamamoto, A. (1988). *Penrose patterns and related structures. I. Superstructure and generalized Penrose patterns. Acta Cryst.* A**44**, 508–516.

Janner, A. (1992). *Decagrammal symmetry of decagonal $Al_{78}Mn_{22}$ quasicrystal. Acta Cryst.* A**48**, 884–901.

Janner, A. & Janssen, T. (1979). *Superspace groups. Physica A*, **99**, 47–76.

Janner, A. & Janssen, T. (1980*a*). *Symmetry of incommensurate crystal phases. I. Commensurate basic structures. Acta Cryst.* A**36**, 399–408.

Janner, A. & Janssen, T. (1980*b*). *Symmetry of incommensurate crystal phases. II. Incommensurate basic structures. Acta Cryst.* A**36**, 408–415.

Janner, A., Janssen, T. & de Wolff, P. M. (1983*a*). *Bravais classes for incommensurate crystal phases. Acta Cryst.* A**39**, 658–666.

Janner, A., Janssen, T. & de Wolff, P. M. (1983*b*). *Determination of the Bravais class for a number of incommensurate crystals. Acta Cryst.* A**39**, 671–678.

Janot, Chr. (1994). *Quasicrystals. A Primer.* Oxford: Clarendon Press.

Janssen, T. (1986). *Crystallography of quasicrystals. Acta Cryst.* A**42**, 261–271.

Janssen, T. (1988). *Aperiodic crystals: a contradictio in terminis? Phys. Rep.* **168**, 55–113.

Janssen, T. (1995). *From quasiperiodic to more complex systems.* In *Beyond Quasicrystals*, edited by F. Axel & D. Gratias, pp. 75–140. Les Ulis: Les Editions de Physique and Berlin: Springer-Verlag.

Janssen, T., Janner, A., Looijenga-Vos, A. & de Wolff, P. M. (2004). *Incommensurate and commensurate modulated crystal structures.* In *International Tables for Crystallography*, Vol. C, edited by E. Prince, ch. 9.8. Dordrecht: Kluwer Academic Publishers.

Jaric, M. V. (1986). *Diffraction from quasicrystals: geometric structure factor. Phys. Rev. B*, **34**, 4685–4698.

Kalning, M., Kek, S., Burandt, B., Press, W. & Steurer, W. (1994). *Examination of a multiple twinned periodic approximant of the decagonal phase $Al_{70}Co_{15}Ni_{15}$. J. Phys. Condens. Matter*, **6**, 6177–6187.

Kato, K. (1990). *Strukturverfeinerung des Kompositkristalls im mehrdimensionalen Raum. Acta Cryst.* B**46**, 39–44.

Kelton, K. F. (1995). *Quasicrystals and related structures.* In *Intermetallic Compounds. Principles and Practice*, Vol. 1, edited by J. H. Westbrook & R. L. Fleischer, pp. 453–491. Chichester: John Wiley & Sons.

Korekawa, M. (1967). *Theorie der Satellitenreflexe.* Habilitation thesis, University of Munich.

Lancon, F., Billard, L., Burkov, S. & DeBoissieu, M. (1994). *On choosing a proper basis for determining structures of quasicrystals. J. Phys. I France*, **4**, 283–301.

Levine, D. & Steinhardt, P. J. (1986). *Quasicrystals. I. Definition and structure. Phys. Rev. B*, **34**, 596–616.

Levitov, L. S. & Rhyner, J. (1988). *Crystallography of quasicrystals; application to icosahedral symmetry. J. Phys. France*, **49**, 1835–1849.

Luck, J. M., Godréche, C., Janner, A. & Janssen, T. (1993). *The nature of the atomic surfaces of quasiperiodic self-similar structures. J. Phys. A Math. Gen.* **26**, 1951–1999.

Paciorek, W. A. & Chapuis, G. (1994). *Generalized Bessel functions in incommensurate structure analysis. Acta Cryst.* A**50**, 194–203.

Pavlovitch, A. & Kleman, M. (1987). *Generalized 2D Penrose tilings: structural properties. J. Phys. A Math. Gen.* **20**, 687–702.

Penrose, R. (1974). *The role of aesthetics in pure and applied mathematical research. Bull. Math. Appl.* **10**, 266–271.

Penrose, R. (1979). *Pentaplexity. A class of non-periodic tilings of the plane. Math. Intell.* **2**, 32–37.

Petricek, V., Maly, K. & Cisarova, I. (1991). *The computing system 'JANA'.* In *Methods of Structural Analysis of Modulated Structures and Quasicrystals*, edited by J. M. Pérez-Mato, F. J. Zuniga & G. Madriaga, pp. 262–267. Singapore: World Scientific.

Petricek, V., Maly, K., Coppens, P., Bu, X., Cisarova, I. & Frost-Jensen, A. (1991). *The description and analysis of composite crystals. Acta Cryst.* A**47**, 210–216.

Rabson, D. A., Mermin, N. D., Rokhsar, D. S. & Wright, D. C. (1991). *The space groups of axial crystals and quasicrystals. Rev. Mod. Phys.* **63**, 699–733.

Rokhsar, D. S., Wright, D. C. & Mermin, N. D. (1988). *Scale equivalence of quasicrystallographic space groups. Phys. Rev. B*, **37**, 8145–8149.

Senechal, M. (1995). *Quasicrystals and Geometry*. Cambridge University Press.

Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. (1984). *Metallic phase with long-range orientational order and no translational symmetry. Phys. Rev. Lett.* **53**, 1951–1953.

Smaalen, S. van (1991). *Symmetry of composite crystals. Phys. Rev. B*, **43**, 11330–11341.

Smaalen, S. van (1992). *Superspace description of incommensurate intergrowth compounds and the application to inorganic misfit layer compounds. Mater. Sci. Forum*, **100 & 101**, 173–222.

Smaalen, S. van (1995). *Incommensurate crystal structures. Crystallogr. Rev.* **4**, 79–202.

Socolar, J. E. S. & Steinhardt, P. J. (1986). *Quasicrystals. II. Unit-cell configurations. Phys. Rev. B*, **34**, 617–647.

Steurer, W. (1990). *The structure of quasicrystals. Z. Kristallogr.* **190**, 179–234.

Steurer, W. (1995). *Experimental aspects of the structure analysis of aperiodic materials.* In *Beyond Quasicrystals*, edited by F. Axel & D. Gratias, pp. 203–228. Les Ulis: Les Editions de Physique and Berlin: Springer-Verlag.

Steurer, W. (1996). *The structure of quasicrystals.* In *Physical Metallurgy*, Vol. I, edited by R. W. Cahn & P. Haasen, pp. 371–411. Amsterdam: Elsevier.

Willis, B. T. M. & Pryor, A. W. (1975). *Thermal Vibrations in Crystallography*. Cambridge University Press.

Wolff, P. M. de (1974). *The pseudo-symmetry of modulated crystal structures. Acta Cryst.* A**30**, 777–785.

Wolff, P. M. de (1977). *Symmetry operations for displacively modulated structures. Acta Cryst.* A**33**, 493–497.

Wolff, P. M. de (1984). *Dualistic interpretation of the symmetry of incommensurate structures. Acta Cryst.* A**40**, 34–42.

Wolff, P. M. de, Janssen, T. & Janner, A. (1981). *The superspace groups for incommensurate crystal structures with a one-dimensional modulation. Acta Cryst.* A**37**, 625–636.

Yamamoto, A. (1982). *Structure factor of modulated crystal structures. Acta Cryst.* A**38**, 87–92.

Yamamoto, A. (1992*a*). *Unified setting and symbols of superspace groups for composite crystals. Acta Cryst.* A**48**, 476–483.

Yamamoto, A. (1992*b*). *Ideal structure of icosahedral Al–Cu–Li quasicrystals. Phys. Rev. B*, **45**, 5217–5227.

Zobetz, E. (1993). *One-dimensional quasilattices: fractally shaped atomic surfaces and homometry. Acta Cryst.* A**49**, 667–676.

International Tables for Crystallography (2010). Vol. B, Chapter 5.1, pp. 626–646.

5.1. Dynamical theory of X-ray diffraction

By A. Authier

5.1.1. Introduction

The first experiment on X-ray diffraction by a crystal was performed by W. Friedrich, P. Knipping and M. von Laue in 1912 and Bragg's law was derived in 1913 (Bragg, 1913). Geometrical and dynamical theories for the intensities of the diffracted X-rays were developed by Darwin (1914*a*,*b*). His dynamical theory took into account the interaction of X-rays with matter by solving recurrence equations that describe the balance of partially transmitted and partially reflected amplitudes at each lattice plane. This is the first form of the dynamical theory of X-ray diffraction. It gives correct expressions for the reflected intensities and was extended to the absorbing-crystal case by Prins (1930). A second form of dynamical theory was introduced by Ewald (1917) as a continuation of his previous work on the diffraction of optical waves by crystals. He took into account the interaction of X-rays with matter by considering the crystal to be a periodic distribution of dipoles which were excited by the incident wave. This theory also gives the correct expressions for the reflected and transmitted intensities, and it introduces the fundamental notion of a wavefield, which is necessary to understand the propagation of X-rays in perfect or deformed crystals. Ewald's theory was later modified by von Laue (1931), who showed that the interaction could be described by solving Maxwell's equations in a medium with a continuous, triply periodic distribution of dielectric susceptibility. It is this form which is most widely used today and which will be presented in this chapter.

The geometrical (or kinematical) theory, on the other hand, considers that each photon is scattered only once and that the interaction of X-rays with matter is so small it can be neglected. It can therefore be assumed that the amplitude incident on every diffraction centre inside the crystal is the same. The total diffracted amplitude is then simply obtained by adding the individual amplitudes diffracted by each diffracting centre, taking into account only the geometrical phase differences between them and neglecting the interaction of the radiation with matter. The result is that the distribution of diffracted amplitudes in reciprocal space is the Fourier transform of the distribution of diffracting centres in physical space. Following von Laue (1960), the expression *geometrical theory* will be used throughout this chapter when referring to these geometrical phase differences.

The first experimentally measured reflected intensities were not in agreement with the theoretical values obtained with the more rigorous dynamical theory, but rather with the simpler geometrical theory. The integrated reflected intensities calculated using geometrical theory are proportional to the square of the structure factor, while the corresponding expressions calculated using dynamical theory for an infinite perfect crystal are proportional to the modulus of the structure factor. The integrated intensity calculated by geometrical theory is also proportional to the volume of the crystal bathed in the incident beam. This is due to the fact that one neglects the decrease of the incident amplitude as it progresses through the crystal and a fraction of it is scattered away. According to geometrical theory, the diffracted intensity would therefore increase to infinity if the volume of the crystal was increased to infinity, which is of course absurd. The theory only works because the strength of the interaction is very weak and if it is applied to very small crystals. How small will be shown quantitatively in Sections 5.1.6.5 and 5.1.7.2. Darwin (1922) showed that it can also be applied to large imperfect crystals. This is done using the model of mosaic crystals (Bragg *et al.*, 1926). For perfect or nearly perfect crystals, dynamical theory should be used. Geometrical theory presents another drawback: it gives no indication as to the phase of the reflected wave. This is due to the fact that it is based on the Fourier transform of the electron density limited by the external shape of the crystal. This is not important when one is only interested in measuring the reflected intensities. For any problem where the phase is important, as is the case for multiple reflections, interference between coherent blocks, standing waves *etc.*, dynamical theory should be used, even for thin or imperfect crystals.

Until the 1940s, the applications of dynamical theory were essentially intensity measurements. From the 1950s to the 1970s, applications were related to the properties (absorption, interference, propagation) of wavefields in perfect or nearly perfect crystals: anomalous transmission, diffraction of spherical waves, interpretation of images on X-ray topographs, accurate measurement of form factors, lattice-parameter mapping. In recent years, they have been concerned mainly with crystal optics, focusing and the design of monochromators for synchrotron radiation [see, for instance, Batterman & Bilderback (1991)], the location of atoms at crystal surfaces and interfaces using the standing-waves method, determination of phases using multiple reflections [for reviews of *n*-beam diffraction, see Weckert & Hümmer (1997) and Chang (2004); for recent determinations of phases, see Chang *et al.* (2002), Mo *et al.* (2002), Weckert *et al.* (2002), Shen & Wang (2003)], characterization of the crystal perfection of epilayers and superlattices by high-resolution diffractometry [see, for instance, Tanner (1990) and Fewster (1993)], *etc.*

Modern developments include the extension of dynamical theory to time-dependent phenomena (Chukhovskii & Förster, 1995; Shastri *et al.*, 2001; Graeff, 2002*a*,*b*, 2004; Malgrange & Graeff, 2003; Sondhauss & Wark, 2003; Adams, 2004) and the study of the influence of the coherence of the source (Yamazaki & Ishikawa, 2002, 2004).

For reviews of dynamical theory, see Zachariasen (1945), von Laue (1960), James (1963), Batterman & Cole (1964), Authier (1970), Kato (1974), Brümmer & Stephanik (1976), Pinsker (1978), Authier *et al.* (1996), Authier & Malgrange (1998), and Authier (2005). Topography is described in Chapter 2.7 of *IT* C (2004), in Tanner (1976) and in Tanner & Bowen (1992). For the use of Bragg-angle measurements for accurate lattice-parameter mapping, see Hart (1981). For online calculations in the case of multiple diffraction, grazing incidence or for strained crystals, see http://sergey.gmca.aps.anl.gov.

A reminder of some basic concepts in electrodynamics is given in Section A5.1.1.1 of the Appendix.

5.1.2. Fundamentals of plane-wave dynamical theory

5.1.2.1. Propagation equation

The wavefunction Ψ associated with an electron or a neutron beam is *scalar* while an electromagnetic wave is a *vector* wave. When propagating in a medium, these waves are solutions of a *propagation equation*. For electrons and neutrons, this is Schrödinger's equation, which can be rewritten as

$$\Delta\Psi + 4\pi^2 k^2(1+\chi)\Psi = 0, \qquad (5.1.2.1)$$

Copyright © 2010 International Union of Crystallography

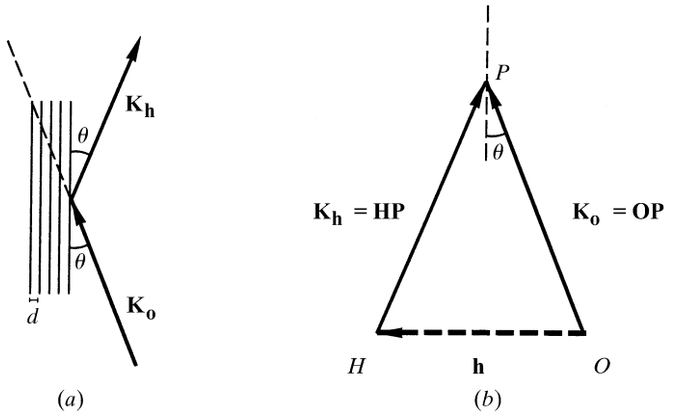


Fig. 5.1.2.1. Bragg reflection. (*a*) Direct space. Bragg reflection of a wave of wavevector $\mathbf{K_o}$ incident on a set of lattice planes of spacing d. The reflected wavevector is $\mathbf{K_h}$. Bragg's law $2d\sin\theta = n\lambda$ can also be written $2d_{hkl}\sin\theta = \lambda$, where $d_{hkl} = d/n = 1/OH = 1/h$ is the inverse of the length of the corresponding reciprocal-lattice vector $\mathbf{OH} = \mathbf{h}$ (see part *b*). (*b*) Reciprocal space. P is the tie point of the wavefield consisting of the incident wave $\mathbf{K_o} = \mathbf{OP}$ and the reflected wave $\mathbf{K_h} = \mathbf{HP}$. Note that the wavevectors are oriented *towards* the tie point.

where $k = 1/\lambda$ is the wavenumber in a vacuum, $\chi = \varphi/W$ (φ is the potential in the crystal and W is the accelerating voltage) in the case of electron diffraction and $\chi = -2mV(\mathbf{r})/h^2k^2$ [$V(\mathbf{r})$ is the Fermi pseudo-potential and h is Planck's constant] in the case of neutron diffraction. The dynamical theory of electron diffraction is treated in Chapter 5.2 [note that a different convention is used in Chapter 5.2 for the scalar wavenumber: $k = 2\pi/\lambda$; compare, for example, equation (5.2.2.1) and its equivalent, equation (5.1.2.1)] and the dynamical theory of neutron diffraction is treated in Chapter 5.3.

In the case of X-rays, the propagation equation is deduced from Maxwell's equations after neglecting the interaction with protons. Following von Laue (1931, 1960), it is assumed that the positive charge of the nuclei is distributed in such a way that the medium is everywhere locally neutral and that there is no current. As a first approximation, magnetic interaction, which is very weak, is not taken into account in this review. The propagation equation is derived in Section A5.1.1.2 of the Appendix. Expressed in terms of the local electric displacement, $\mathbf{D}(\mathbf{r})$, it is given for monochromatic waves by

$$\Delta\mathbf{D}(\mathbf{r}) + \text{ curl curl } \chi\mathbf{D}(\mathbf{r}) + 4\pi^2k^2\mathbf{D}(\mathbf{r}) = 0. \qquad (5.1.2.2)$$

The interaction of X-rays with matter is characterized in equation (5.1.2.2) by the parameter χ, which is the dielectric susceptibility. It is classically related to the electron density $\rho(\mathbf{r})$ by

$$\chi(\mathbf{r}) = -R\lambda^2\rho(\mathbf{r})/\pi, \qquad (5.1.2.3)$$

where $R = 2.81794 \times 10^{-6}$ nm is the classical radius of the electron [see equation (A5.1.1.2) in Section A5.1.1.2 of the Appendix].

The dielectric susceptibility, being proportional to the electron density, is triply periodic in a crystal. It can therefore be expanded in Fourier series:

$$\chi = \sum_{\mathbf{h}} \chi_h \exp(2\pi i\mathbf{h}\cdot\mathbf{r}), \qquad (5.1.2.4)$$

where $\mathbf{h}$ is a reciprocal-lattice vector and the summation is extended over all reciprocal-lattice vectors. The sign convention adopted here for Fourier expansions of periodic functions is the *standard crystallographic* sign convention defined in Section 2.5.2.3. The relative orientations of wavevectors and reciprocal-lattice vectors are defined in Fig. 5.1.2.1, which represents schematically a Bragg reflection in direct and reciprocal space (Figs. 5.1.2.1*a* and 5.1.2.1*b*, respectively).

The coefficients χ_h of the Fourier expansion of the dielectric susceptibility are related to the usual structure factor F_h by

$$\chi_h = -R\lambda^2F_h/(\pi V), \qquad (5.1.2.5)$$

where V is the volume of the unit cell and the structure factor is given by

$$\begin{aligned} F_h &= \sum_j (f_j + f_j' + if_j'')\exp(-M_j - 2\pi i\mathbf{h}\cdot\mathbf{r}_j) \\ &= |F_h|\exp(i\varphi_h). \end{aligned} \qquad (5.1.2.6)$$

f_j is the form factor of atom j, f_j' and f_j'' are the dispersion corrections [see, for instance, *IT* C, Section 4.2.6] and $\exp(-M_j)$ is the Debye–Waller factor. The summation is over all the atoms in the unit cell. The phase φ_h of the structure factor depends of course on the choice of origin of the unit cell. The Fourier coefficients χ_h are dimensionless. Their order of magnitude varies from 10^{-5} to 10^{-7} depending on the wavelength and the structure factor. For example, χ_h is -9.24×10^{-6} for the 220 reflection of silicon for Cu $K\alpha$ radiation.

In an absorbing crystal, absorption is taken into account phenomenologically through the imaginary parts of the index of refraction and of the wavevectors. The dielectric susceptibility is written

$$\chi = \chi_\mathrm{r} + i\chi_\mathrm{i}. \qquad (5.1.2.7)$$

The real and imaginary parts of the susceptibility are triply periodic in a crystalline medium and can be expanded in a Fourier series,

$$\begin{aligned} \chi_\mathrm{r} &= \sum_{\mathbf{h}} \chi_{\mathrm{r}h}\exp(2\pi i\mathbf{h}\cdot\mathbf{r}) \\ \chi_\mathrm{i} &= \sum_{\mathbf{h}} \chi_{\mathrm{i}h}\exp(2\pi i\mathbf{h}\cdot\mathbf{r}), \end{aligned} \qquad (5.1.2.8)$$

where

$$\begin{aligned} \chi_{\mathrm{r}h} &= -R\lambda^2F_{\mathrm{r}h}/(\pi V), \\ \chi_{\mathrm{i}h} &= -R\lambda^2F_{\mathrm{i}h}/(\pi V) \end{aligned} \qquad (5.1.2.9)$$

and

$$\begin{aligned} F_{\mathrm{r}h} &= \sum_j (f_j + f_j')\exp(-M_j - 2\pi i\mathbf{h}\cdot\mathbf{r}_j) \\ &= |F_{\mathrm{r}h}|\exp(i\varphi_{\mathrm{r}h}), \qquad (5.1.2.10a) \\ F_{\mathrm{i}h} &= \sum_j (f_j'')\exp(-M_j - 2\pi i\mathbf{h}\cdot\mathbf{r}_j) \\ &= |F_{\mathrm{i}h}|\exp(i\varphi_{\mathrm{i}h}). \qquad (5.1.2.10b) \end{aligned}$$

It is important to note that

$$F_{\mathrm{r}h}^* = F_{\mathrm{r}\bar{\mathbf{h}}} \text{ and } F_{\mathrm{i}h}^* = F_{\mathrm{i}\bar{\mathbf{h}}} \text{ but that } F_h^* \neq F_{\bar{\mathbf{h}}}, \qquad (5.1.2.11)$$

where f^* is the imaginary conjugate of f.

The *index of refraction* of the medium for X-rays is

$$n = 1 + \chi_{ro}/2 = 1 - R\lambda^2 F_o/(2\pi V), \qquad (5.1.2.12)$$

where F_o/V is the number of electrons per unit volume. This index is very slightly smaller than one. It is for this reason that specular reflection of X-rays takes place at grazing angles. From the value of the critical angle, $(-\chi_{ro})^{1/2}$, the electron density F_o/V of a material can be determined.

The linear absorption coefficient is

$$\mu_o = -2\pi k \chi_{io} = 2R\lambda F_{io}/V. \qquad (5.1.2.13)$$

For example, it is 143.2 cm^{-1} for silicon and Cu $K\alpha$ radiation.

5.1.2.2. Wavefields

The notion of a wavefield, introduced by Ewald (1917), is one of the most fundamental concepts in dynamical theory. It results from the fact that since the propagation equation (5.1.2.2) is a second-order partial differential equation with a periodic interaction coefficient, its solution has the same periodicity,

$$\mathbf{D} = \exp(-2\pi i \mathbf{K_o} \cdot \mathbf{r}) \sum_{\mathbf{h}} \mathbf{D_h} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}), \qquad (5.1.2.14)$$

where the summation is over all reciprocal-lattice vectors **h**. Equation (5.1.2.14) can also be written

$$\mathbf{D} = \sum_{\mathbf{h}} \mathbf{D_h} \exp(-2\pi i \mathbf{K_h} \cdot \mathbf{r}), \qquad (5.1.2.15)$$

where

$$\mathbf{K_h} = \mathbf{K_o} - \mathbf{h}. \qquad (5.1.2.16)$$

Expression (5.1.2.15) shows that the solution of the propagation equation can be interpreted as an infinite sum of plane waves with amplitudes $\mathbf{D_h}$ and wavevectors $\mathbf{K_h}$. This sum is a *wavefield*, or *Ewald wave*. The same expression is used to describe the propagation of any wave in a periodic medium, such as phonons or electrons in a solid. Expression (5.1.2.14) was later called a *Bloch wave* by solid-state physicists.

The wavevectors in a wavefield are deduced from one another by translations of the reciprocal lattice [expression (5.1.2.16)]. They can be represented geometrically as shown in Fig. 5.1.2.1(*b*). The wavevectors $\mathbf{K_o} = \mathbf{OP}$; $\mathbf{K_h} = \mathbf{HP}$ are drawn *away* from reciprocal-lattice points. Their common extremity, *P*, called the *tie point* by Ewald, characterizes the wavefield.

In an absorbing crystal, wavevectors have an imaginary part,

$$\mathbf{K_o} = \mathbf{K_{or}} + i\mathbf{K_{oi}}; \quad \mathbf{K_h} = \mathbf{K_{hr}} + i\mathbf{K_{hi}},$$

and (5.1.2.16) shows that all wavevectors have the same imaginary part,

$$\mathbf{K_{oi}} = \mathbf{K_{hi}}, \qquad (5.1.2.17)$$

and therefore undergo the same absorption. This is one of the most important properties of wavefields.

The first experimental evidence of the physical existence of Ewald's wavefields is to be found in the light–dark structure of Kossel lines, observed by Borrmann (1936) and explained by von Laue (1937) using the properties of the standing waves formed by the wavefields inside the crystal. It was followed by the discovery of anomalous absorption (Borrmann, 1941, 1950) and the theoretical determination of the wavefield trajectories (von Laue, 1952; Kato, 1952, 1958; Ewald, 1958), which was confirmed by the experimental observation of spherical wave *Pendellösung* (Kato & Lang, 1959) and by that of the double refraction of X-rays, predicted by Borrmann (1955) and observed by Authier (1960, 1961).

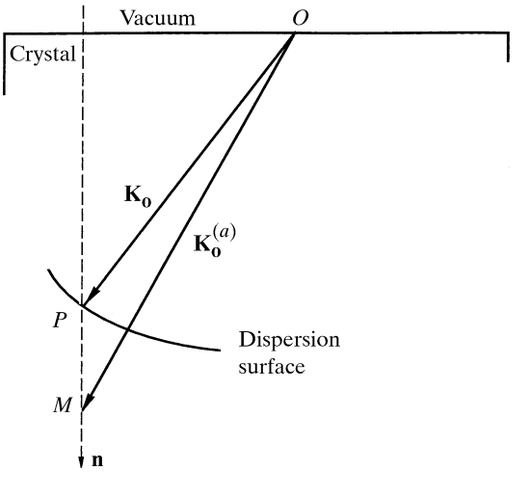


Fig. 5.1.2.2. Boundary condition for wavevectors at the entrance surface of the crystal.

5.1.2.3. Boundary conditions at the entrance surface

The choice of the '**o**' component of expansion (5.1.2.15) is arbitrary in an infinite medium. In a semi-infinite medium where the waves are created at the interface with a vacuum or air by an incident plane wave with wavevector $\mathbf{K_o^{(a)}}$ (using von Laue's notation), the choice of $\mathbf{K_o}$ is determined by the boundary conditions.

This condition for wavevectors at an interface demands that their tangential components should be continuous across the boundary, in agreement with Descartes–Snell's law. This condition is satisfied when the difference between the wavevectors on each side of the interface is parallel to the normal to the interface. This is shown geometrically in Fig. 5.1.2.2 and formally in (5.1.2.18):

$$\mathbf{K_o} - \mathbf{K_o^{(a)}} = \mathbf{OP} - \mathbf{OM} = \overline{MP} \cdot \mathbf{n}, \qquad (5.1.2.18)$$

where **n** is a unit vector normal to the crystal surface, oriented towards the inside of the crystal.

There is no absorption in a vacuum and the incident wavevector $\mathbf{K_o^{(a)}}$ is real. Equation (5.1.2.18) shows that it is the component normal to the interface of wavevector $\mathbf{K_o}$ which has an imaginary part,

$$\mathbf{K_{oi}} = \mathcal{I}(\overline{MP}) \cdot \mathbf{n} = -\mu \mathbf{n}/(4\pi\gamma_o), \qquad (5.1.2.19)$$

where $\mathcal{I}(f)$ is the imaginary part of f, $\gamma_o = \cos(\mathbf{n} \cdot \mathbf{s_o})$ and $\mathbf{s_o}$ is a unit vector in the incident direction. When there is more than one wave in the wavefield, the effective absorption coefficient μ can differ significantly from the normal value, μ_o, given by (5.1.2.13) – see Section 5.1.5.

5.1.2.4. Fundamental equations of dynamical theory

In order to obtain the solution of dynamical theory, one inserts expansions (5.1.2.15) and (5.1.2.4) into the propagation equation (5.1.2.2). This leads to an equation with an infinite sum of terms. It is shown to be equivalent to an infinite system of linear equations which are the *fundamental equations* of dynamical

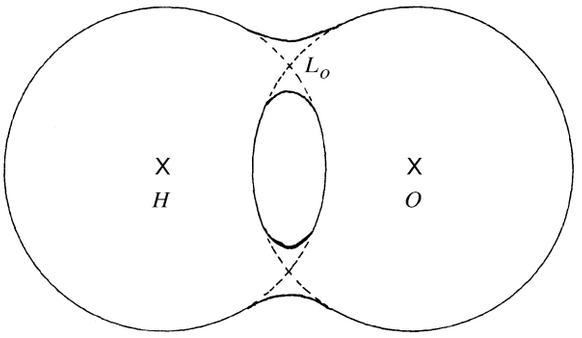


Fig. 5.1.2.3. Intersection of the dispersion surface with the plane of incidence. The dispersion surface is a connecting surface between the two spheres centred at reciprocal-lattice points O and H and with radius nk. L_o is the Lorentz point.

theory. Only those terms in (5.1.2.15) whose wavevector magnitudes K_h are very close to the vacuum value, k, have a non-negligible amplitude. These wavevectors are associated with reciprocal-lattice points that lie very close to the Ewald sphere. Far from any Bragg reflection, their number is equal to 1 and a single plane wave propagates through the medium. In general, for X-rays, there are only two reciprocal-lattice points on the Ewald sphere. This is the so-called *two-beam* case to which this treatment is limited. There are, however, many instances where several reciprocal-lattice points lie simultaneously on the Ewald sphere. This corresponds to the *many-beam* case which has interesting applications for the determination of phases of reflections [see, for instance, Chang (1987, 2004), Hümmer & Weckert (1995) and Weckert & Hümmer (1997)]. On the other hand, for electrons, there are in general many reciprocal-lattice points close to the Ewald sphere and many wavefields are excited simultaneously (see Chapter 5.2).

In the two-beam case, for reflections that are not highly asymmetric and for Bragg angles that are not close to $\pi/2$, the fundamental equations of dynamical theory reduce to

$$\begin{aligned} 2X_oD_o - kC\chi_{\bar{h}}D_h &= 0 \\ -kC\chi_hD_o + 2X_hD_h &= 0, \end{aligned} \qquad (5.1.2.20)$$

where $C = 1$ if $\mathbf{D_h}$ is normal to the $\mathbf{K_o}$, $\mathbf{K_h}$ plane and $C = \cos 2\theta$ if $\mathbf{D_h}$ lies in the plane; this is due to the fact that the amplitude with which electromagnetic radiation is scattered is proportional to the sine of the angle between the direction of the electric vector of the incident radiation and the direction of scattering (see, for instance, *IT* C, Section 6.2.2). The polarization of an electromagnetic wave is classically related to the orientation of the electric vector; in dynamical theory it is that of the electric displacement which is considered (see Section A5.1.1.3 of the Appendix).

The system (5.1.2.20) is therefore a system of four equations which admits four solutions, two for each direction of polarization. In the *non-absorbing* case, to a very good approximation,

$$\begin{aligned} X_o &= K_o - nk, \\ X_h &= K_h - nk. \end{aligned} \qquad (5.1.2.21)$$

In the case of an *absorbing crystal*, X_o and K_h are complex. Equation (5.1.5.2) gives the full expression for X_o.

5.1.2.5. Dispersion surface

The fundamental equations (5.1.2.20) of dynamical theory are a set of linear homogeneous equations whose unknowns are the amplitudes of the various waves which make up a wavefield. For the solution to be nontrivial, the determinant of the set must be set equal to zero. This provides a secular equation relating the magnitudes of the wavevectors of a given wavefield. This equation is that of the locus of the tie points of all the wavefields that may propagate in the crystal with a given frequency. This locus is called the *dispersion surface*. It is a constant-energy surface and is the equivalent of the index surface in optics. It is the X-ray analogue of the constant-energy surfaces known as Fermi surfaces in the electron band theory of solids.

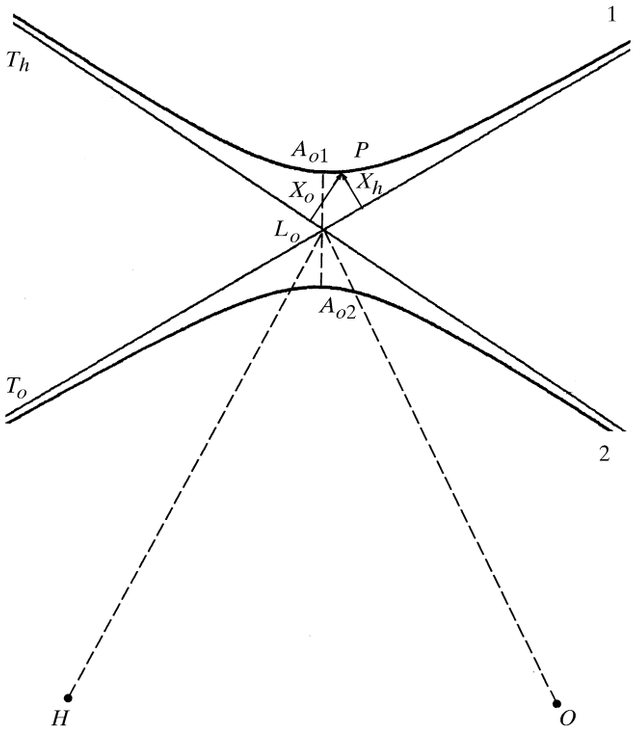


Fig. 5.1.2.4. Intersection of the dispersion surface with the plane of incidence shown in greater detail. The Lorentz point L_o is far away from the nodes O and H of the reciprocal lattice: $OL_o = HL_o = 1/\lambda$ is about 10^5 to 10^6 times larger than the diameter $A_{o1}A_{o2}$ of the dispersion surface.

In the two-beam case, the dispersion surface is a surface of revolution around the diffraction vector **OH**. It is made from two spheres and a connecting surface between them. The two spheres are centred at O and H and have the same radius, nk. Fig. 5.1.2.3 shows the intersection of the dispersion surface with a plane passing through **OH**. When the tie point lies on one of the two spheres, far from their intersection, only one wavefield propagates inside the crystal. When it lies on the connecting surface, two waves are excited simultaneously. The equation of this surface is obtained by equating to zero the determinant of system (5.1.2.20):

$$X_oX_h = k^2C^2\chi_h\chi_{\bar{h}}/4. \qquad (5.1.2.22)$$

Equations (5.1.2.21) show that, in the zero-absorption case, X_o and X_h are to be interpreted as the distances of the tie point P from the spheres centred at O and H, respectively. From (5.1.2.20) it can be seen that they are of the order of the vacuum wavenumber times the Fourier coefficient of the dielectric susceptibility, that is five or six orders of magnitude smaller than k. The two spheres can therefore be replaced by their tangential planes. Equation (5.1.2.22) shows that the product of the distances of the tie point from these planes is constant. The intersection of the dispersion surface with the plane passing through **OH** is therefore a hyperbola (Fig. 5.1.2.4) whose diameter [using (5.1.2.5) and (5.1.2.22)] is

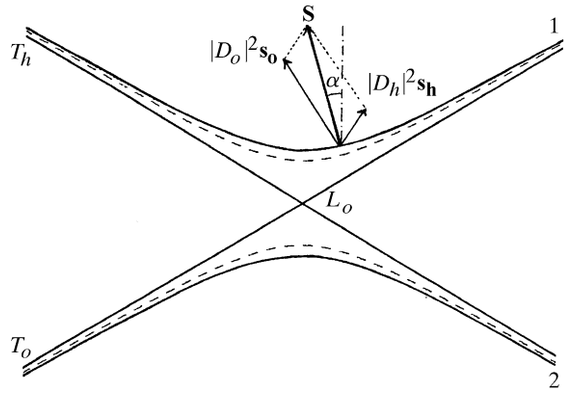


Fig. 5.1.2.5. Dispersion surface for the two states of polarization. Solid curve: polarization normal to the plane of incidence ($C = 1$); broken curve: polarization parallel to the plane of incidence ($C = \cos 2\theta$). The direction of propagation of the energy of the wavefields is along the Poynting vector, **S**, normal to the dispersion surface.

$$\overline{A_{o2}A_{o1}} = |C|R\lambda(F_hF_{\bar{h}})^{1/2}/(\pi V\cos\theta). \qquad (5.1.2.23)$$

It can be noted that the larger the diameter of the dispersion surface, the larger the structure factor, that is, the stronger the interaction of the waves with the matter. When the polarization is parallel to the plane of incidence ($C = \cos 2\theta$), the interaction is weaker.

The asymptotes T_o and T_h to the hyperbola are tangents to the circles centred at O and H, respectively. Their intersection, L_o, is called the *Lorentz point* (Fig. 5.1.2.4).

A wavefield propagating in the crystal is characterized by a tie point P on the dispersion surface and two waves with wavevectors $\mathbf{K_o} = \mathbf{OP}$ and $\mathbf{K_h} = \mathbf{HP}$, respectively. The ratio, ξ, of their amplitudes D_h and D_o is given by means of (5.1.2.20):

$$\xi = \frac{D_h}{D_o} = \frac{2X_o}{kC\chi_{\bar{h}}} = \frac{-2\pi VX_o}{R\lambda CF_{\bar{h}}}. \qquad (5.1.2.24)$$

The hyperbola has two branches, 1 and 2, for each direction of polarization, that is, for $C = 1$ or $\cos 2\theta$ (Fig. 5.1.2.5). Branch 2 is the one situated on the same side of the asymptotes as the reciprocal-lattice points O and H. Given the orientation of the wavevectors, which has been chosen away from the reciprocal-lattice points (Fig. 5.1.2.1*b*), the coordinates of the tie point, X_o and X_h, are positive for branch 1 and negative for branch 2. The phase of ξ is therefore equal to $\pi + \varphi_h$ and to φ_h for the two branches, respectively, where φ_h is the phase of the structure factor [equation (5.1.2.6)]. This difference of π between the two branches has important consequences for the properties of the wavefields.

As mentioned above, owing to absorption, wavevectors are actually complex and so is the dispersion surface (see Authier, 2008).

5.1.2.6. Propagation direction

The energy of all the waves in a given wavefield propagates in a common direction, which is given by that of the Poynting vector (von Laue, 1952) [see Section A5.1.1.4, equation (A5.1.1.8) of the Appendix]. It can be shown that, averaged over time and the unit cell, the Poynting vector of a wavefield is

$$\mathbf{S} = (c/\varepsilon_0)\exp(4\pi\mathbf{K_{oi}}\cdot\mathbf{r})\big[|D_o|^2\mathbf{s_o} + |D_h|^2\mathbf{s_h}\big], \qquad (5.1.2.25)$$

where $\mathbf{s_o}$ and $\mathbf{s_h}$ are unit vectors in the $\mathbf{K_o}$ and $\mathbf{K_h}$ directions, respectively, c is the velocity of light and ε_0 is the dielectric permittivity of a vacuum.

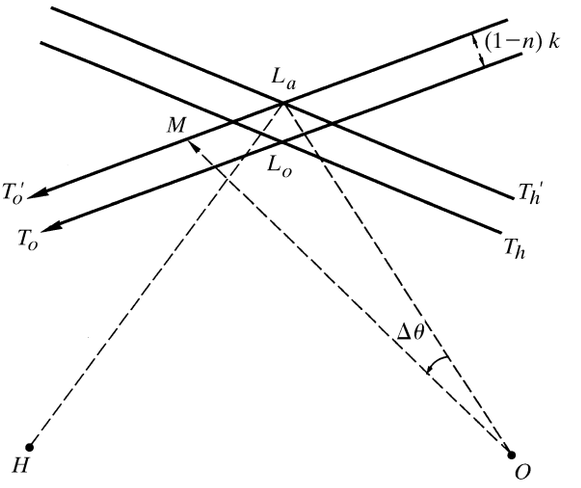


Fig. 5.1.3.1. Departure from Bragg's law of an incident wave.

From (5.1.2.25) and equation (5.1.2.22) of the dispersion surface, it can be shown that the propagation direction of the wavefield lies along the normal to the dispersion surface at the tie point (Fig. 5.1.2.5). This result was first shown by Kato (1952) in the two-beam case for electron diffraction and generalized by him to the n-beam case for X-rays (Kato, 1958). It is also obtained by considering the group velocity of the wavefield (Ewald, 1958; Wagner, 1959). The angle α between the propagation direction and the lattice planes is given by

$$\tan\alpha = \big[(1 - |\xi|^2)/(1 + |\xi|^2)\big]\tan\theta. \qquad (5.1.2.26)$$

It should be noted that the propagation direction varies between $\mathbf{K_o}$ and $\mathbf{K_h}$ for both branches of the dispersion surface.

5.1.3. Solutions of plane-wave dynamical theory

5.1.3.1. Departure from Bragg's law of the incident wave

The wavefields excited in the crystal by the incident wave are determined by applying the boundary condition mentioned above for the continuity of the tangential component of the wavevectors (Section 5.1.2.3). Waves propagating in a vacuum have wavenumber $k = 1/\lambda$. Depending on whether they propagate in the incident or in the reflected direction, the common extremity, M, of their wavevectors

$$\mathbf{OM} = \mathbf{K_o^{(a)}} \text{ and } \mathbf{HM} = \mathbf{K_h^{(a)}}$$

lies on spheres of radius k and centred at O and H, respectively. The intersections of these spheres with the plane of incidence are two circles which can be approximated by their tangents T_o' and T_h' at their intersection point, L_a, or *Laue point* (Fig. 5.1.3.1).

Bragg's condition is exactly satisfied according to the geometrical theory of diffraction when M lies at L_a. The departure $\Delta\theta$ from Bragg's incidence of an incident wave is defined as the angle between the corresponding wavevectors **OM** and $\mathbf{OL}_a$. As $\Delta\theta$ is very small compared to the Bragg angle in the general case of X-rays or neutrons, one may write

$$\begin{aligned}\mathbf{K_o^{(a)}} &= \mathbf{OM} = \mathbf{OL}_a + \mathbf{L}_a\mathbf{M},\\ \Delta\theta &= \overline{L_aM}/k.\end{aligned} \qquad (5.1.3.1)$$

The tangent T_o' is oriented in such a way that $\Delta\theta$ is negative when the angle of incidence is smaller than the Bragg angle.

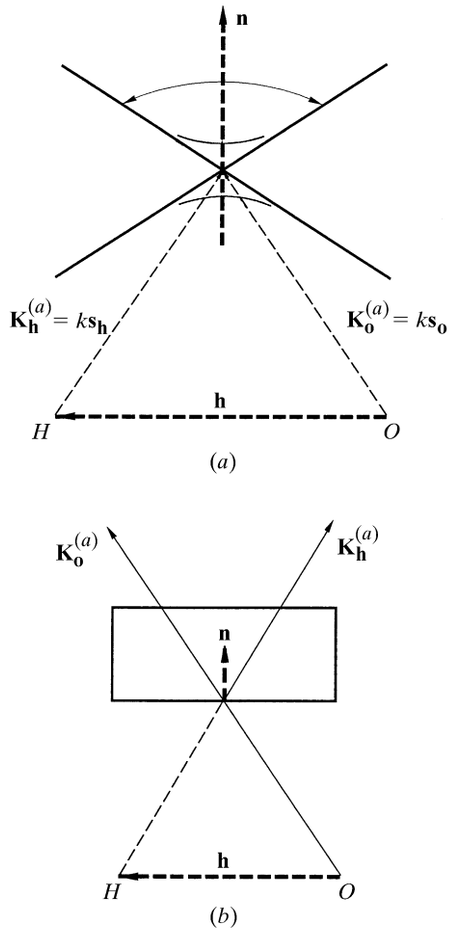


Fig. 5.1.3.2. Transmission, or Laue, geometry. (*a*) Reciprocal space; (*b*) direct space.

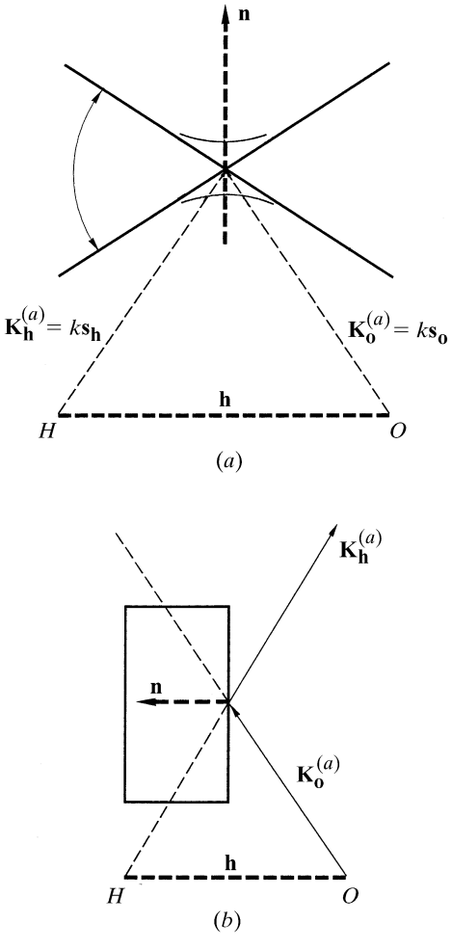


Fig. 5.1.3.3. Reflection, or Bragg, geometry. (*a*) Reciprocal space; (*b*) direct space.

5.1.3.2. Transmission and reflection geometries

The boundary condition for the continuity of the tangential component of the wavevectors is applied by drawing from M a line, **Mz**, parallel to the normal **n** to the crystal surface. The tie points of the wavefields excited in the crystal by the incident wave are at the intersections of this line with the dispersion surface. Two different situations may occur:

(*a*) *Transmission, or Laue case* (Fig. 5.1.3.2). The normal to the crystal surface drawn from M intersects *both* branches of the dispersion surface (Fig. 5.1.3.2*a*). The reflected wave is then directed towards the *inside* of the crystal (Fig. 5.1.3.2*b*). Let γ_o and γ_h be the cosines of the angles between the normal to the crystal surface, **n**, and the incident and reflected directions, respectively:

$$\gamma_o = \cos(\mathbf{n}, \mathbf{s_o}); \quad \gamma_h = \cos(\mathbf{n}, \mathbf{s_h}). \tag{5.1.3.2}$$

It will be noted that they are both positive, as is their ratio,

$$\gamma = \gamma_h/\gamma_o. \tag{5.1.3.3}$$

This is the *asymmetry ratio*, which is very important since the width of the rocking curve is proportional to its square root [equation (5.1.3.6)].

(*b*) *Reflection, or Bragg case* (Fig. 5.1.3.3). In this case there are three possible situations: the normal to the crystal surface drawn from M intersects either branch 1 or branch 2 of the dispersion surface, or the intersection points are imaginary (Fig. 5.1.3.3*a*). The reflected wave is directed towards the *outside* of the crystal (Fig. 5.1.3.3*b*). The cosines defined by (5.1.3.2) are now positive for γ_o and negative for γ_h. The asymmetry factor is therefore also negative.

5.1.3.3. Middle of the reflection domain

It will be apparent from the equations given later that the incident wavevector corresponding to the middle of the reflection domain is, in both cases, **OI**, where I is the intersection of the normal to the crystal surface drawn from the Lorentz point, L_o, with T_o' (Figs. 5.1.3.4 and 5.1.3.5), while, according to Bragg's law, it should be $\mathbf{OL}_a$. The angle $\Delta\theta$ between the incident wavevectors $\mathbf{OL}_a$ and **OI**, corresponding to the middle of the reflecting domain according to the geometrical and dynamical theories, respectively, is

$$\Delta\theta_o = \overline{L_a I}/k = R\lambda^2 F_o(1-\gamma)/(2\pi V \sin 2\theta). \tag{5.1.3.4}$$

In the Bragg case, the asymmetry ratio γ is negative and $\Delta\theta_o$ is never equal to zero. This difference in Bragg angle between the two theories is due to the refraction effect, which is neglected in geometrical theory. In the Laue case, $\Delta\theta_o$ is equal to zero for symmetric reflections ($\gamma = 1$).

5.1.3.4. Deviation parameter

The solutions of dynamical theory are best described by introducing a reduced parameter called the *deviation parameter*,

$$\eta = (\Delta\theta - \Delta\theta_o)/\delta, \tag{5.1.3.5}$$

where

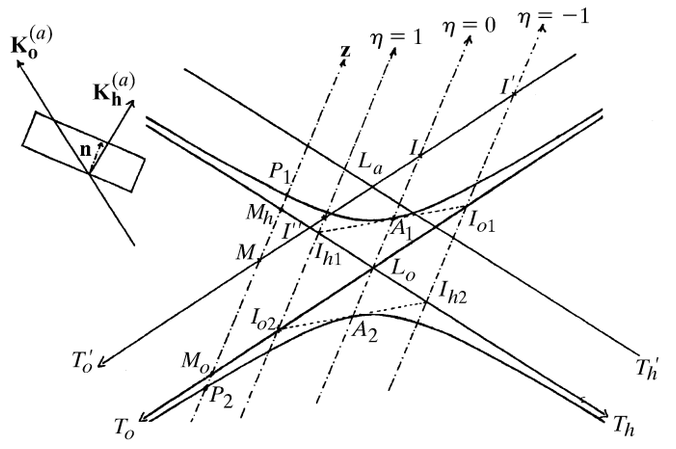


Fig. 5.1.3.4. Boundary conditions at the entrance surface for transmission geometry.

$$\delta = R\lambda^2|C|(|\gamma|F_hF_{\bar{h}})^{1/2}/(\pi V \sin 2\theta), \tag{5.1.3.6}$$

whose real part is equal to the half width of the rocking curve (Sections 5.1.6 and 5.1.7). The width 2δ of the rocking curve is sometimes called the *Darwin width*.

The definition (5.1.3.5) of the deviation parameter is independent of the geometrical situation (reflection or transmission case); this is not followed by some authors. The present convention has the advantage of being quite general.

In an absorbing crystal, η, $\Delta\theta_o$ and δ are complex, and it is the real part, $\Delta\theta_{or}$, of $\Delta\theta_o$ which has the geometrical interpretation given in Section 5.1.3.3. One obtains

$$\begin{aligned}
&\eta = \eta_r + i\eta_i \\
&\eta_r = (\Delta\theta - \Delta\theta_{or})/\delta_r; \;\; \eta_i = A\eta_r + B \\
&A = -\tan\beta \\
&B = \left\{\chi_{io}/\left[|C|(|\chi_h\chi_{\bar{h}}|)^{1/2}\cos\beta\right]\right\}(1-\gamma)/2(|\gamma|)^{1/2},
\end{aligned} \tag{5.1.3.7}$$

where β is the phase angle of $(\chi_h\chi_{\bar{h}})^{1/2}$ [or that of $(F_hF_{\bar{h}})^{1/2}$].

5.1.3.5. Pendellösung and extinction distances

Let

$$\Lambda_o = \pi V(\gamma_o|\gamma_h|)^{1/2}/[R\lambda|C|(F_hF_{\bar{h}})^{1/2}]. \tag{5.1.3.8}$$

This length plays a very important role in the dynamical theory of diffraction by both perfect and deformed crystals. For example, it is 15.3 μm for the 220 reflection of silicon, with Mo $K\alpha$ radiation and a symmetric reflection.

In *transmission* geometry, it gives the period of the interference between the two excited wavefields which constitutes the *Pendellösung* effect first described by Ewald (1917) (see Section 5.1.6.3); Λ_o in this case is called the *Pendellösung distance*, denoted Λ_L hereafter. Its geometrical interpretation, in the zero-absorption case, is the inverse of the diameter A_2A_1 of the dispersion surface in a direction defined by the cosines γ_h and γ_o with respect to the reflected and incident directions, respectively (Fig. 5.1.3.4). It reduces to the inverse of $A_{o2}A_{o1}$ (5.1.2.23) in the symmetric case.

In *reflection* geometry, it gives the absorption distance in the total-reflection domain and is called the *extinction* distance, denoted Λ_B (see Section 5.1.7.1). Its geometrical interpretation in the zero-absorption case is the inverse of the length $I_{o1}I_{h1} = I_{h2}I_{o2}$, Fig. 5.1.3.5.

In a *deformed* crystal, if distortions are of the order of the width of the rocking curve over a distance Λ_o, the crystal is considered to be slightly deformed, and ray theory (Penning & Polder, 1961; Kato, 1963, 1964*a*,*b*) can be used to describe the propagation of wavefields. If the distortions are larger, new wavefields may be generated by interbranch scattering (Authier & Balibar, 1970) and generalized dynamical diffraction theory such as that developed by Takagi (1962, 1969) should be used.

Using (5.1.3.8), expressions (5.1.3.5) and (5.1.3.6) can be rewritten in the very useful form:

$$\begin{aligned}
&\eta = (\Delta\theta - \Delta\theta_o)\Lambda_o \sin 2\theta/(\lambda|\gamma_h|), \\
&\delta = \lambda|\gamma_h|/(\Lambda_o \sin 2\theta).
\end{aligned} \tag{5.1.3.9}$$

The order of magnitude of the Darwin width 2δ ranges from a fraction of a second of an arc to ten or more seconds, and

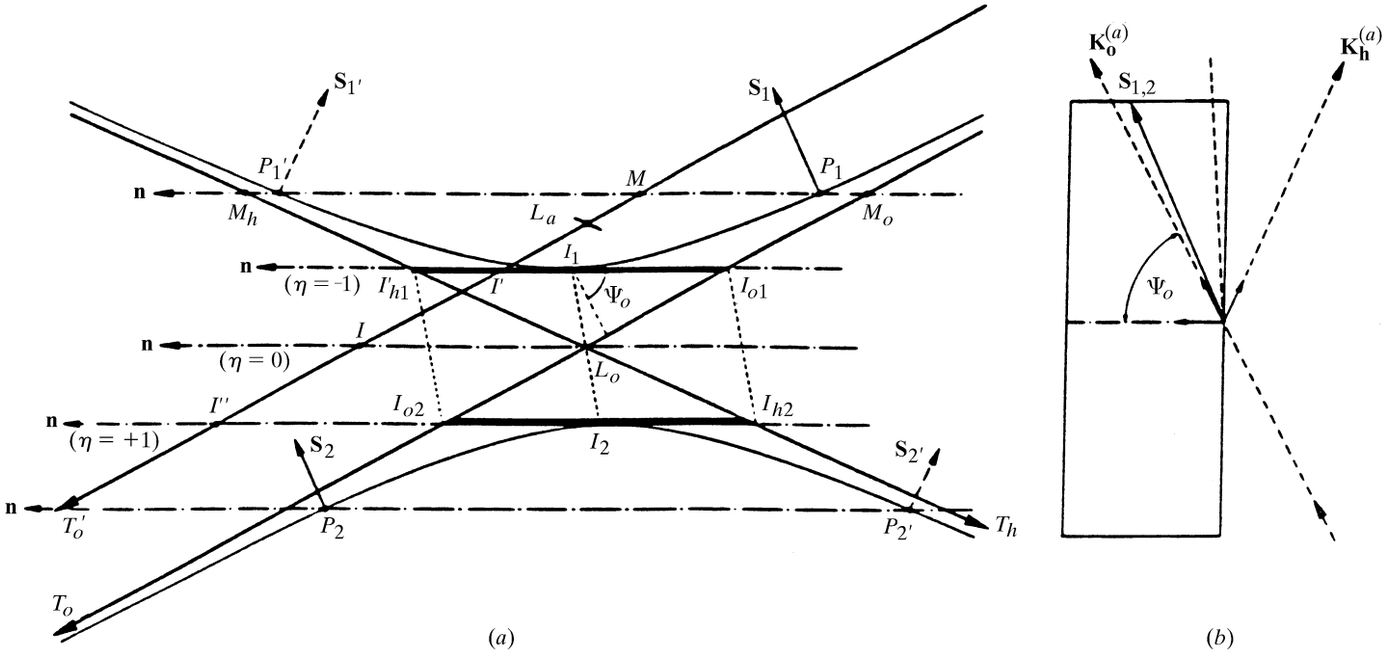


Fig. 5.1.3.5. Boundary conditions at the entrance surface for reflection geometry. (*a*) Reciprocal space; (*b*) direct space.

increases with increasing wavelength and increasing structure factor. For example, for the 220 reflection of silicon and Cu $K\alpha$ radiation, it is 5.2 seconds.

5.1.3.6. Solution of the dynamical theory

The coordinates of the tie points excited by the incident wave are obtained by looking for the intersection of the dispersion surface, (5.1.2.22), with the normal **Mz** to the crystal surface (Figs. 5.1.3.4 and 5.1.3.5). The ratio ξ of the amplitudes of the waves of the corresponding wavefields is related to these coordinates by (5.1.2.24) and is found to be

$$\begin{aligned}\xi_j &= D_{hj}/D_{oj} \\ &= -\mathrm{S}(C)\mathrm{S}(\gamma_h)[(F_h F_{\bar{h}})^{1/2}/F_{\bar{h}}] \\ &\quad \times \left\{\eta \pm \left[\eta^2 + \mathrm{S}(\gamma_h)\right]^{1/2}\right\}/(|\gamma|)^{1/2},\end{aligned} \tag{5.1.3.10}$$

where the plus sign corresponds to a tie point on branch 1 ($j = 1$) and the minus sign to a tie point on branch 2 ($j = 2$), and $\mathrm{S}(\gamma_h)$ is the sign of γ_h (+1 in transmission geometry, -1 in reflection geometry).

5.1.3.7. Geometrical interpretation of the solution in the zero-absorption case

5.1.3.7.1. Transmission geometry

In this case (Fig. 5.1.3.4) $\mathrm{S}(\gamma_h)$ is +1 and (5.1.3.10) may be written

$$\xi_j = -\mathrm{S}(C)\left[\eta \pm (\eta^2 + 1)^{1/2}\right]/\gamma^{1/2}. \tag{5.1.3.11}$$

Let A_1 and A_2 be the intersections of the normal to the crystal surface drawn from the Lorentz point L_o with the two branches of the dispersion surface (Fig. 5.1.3.4). From Sections 5.1.3.3 and 5.1.3.4, they are the tie points excited for $\eta = 0$ and correspond to the middle of the reflection domain. Let us further consider the tangents to the dispersion surface at A_1 and A_2 and let I_{o1}, I_{o2} and I_{h1}, I_{h2} be their intersections with T_o and T_h, respectively. It can be shown that $\overline{I_{o1}I_{h2}}$ and $\overline{I_{o2}I_{h1}}$ intersect the dispersion surface at the tie points excited for $\eta = -1$ and $\eta = +1$, respectively, and that the *Pendellösung* distance $\Lambda_\mathrm{L} = 1/\overline{A_2A_1}$, the width of the rocking curve $2\delta = \overline{I_{o1}I_{o2}}/k$ and the deviation parameter $\eta = \overline{M_oM_h}/\overline{A_2A_1}$, where M_o and M_h are the intersections of the normal to the crystal surface drawn from the extremity of any incident wavevector **OM** with T_o and T_h, respectively.

5.1.3.7.2. Reflection geometry

In this case (Fig. 5.1.3.5) $\mathrm{S}(\gamma_h)$ is now -1 and (5.1.3.10) may be written

$$\xi_j = \mathrm{S}(C)\left[\eta \pm (\eta^2 - 1)^{1/2}\right]/|\gamma|^{1/2}. \tag{5.1.3.12}$$

Now, let I_1 and I_2 be the points of the dispersion surface where the tangent is parallel to the normal to the crystal surface, and further let I_{o1}, I_{h1}, I' and I_{o2}, I_{h2}, I'' be the intersections of these two tangents with T_o, T_h and T'_o, respectively. For an incident wave of wavevector **OM** where M lies between I' and I'', the normal to the crystal surface drawn from M has no real intersection with the dispersion surface and $\overline{I'I''}$ defines the total-reflection domain. The tie points I_1 and I_2 correspond to $\eta = -1$ and $\eta = +1$, respectively, the extinction distance $\Lambda_\mathrm{L} = 1/\overline{I_{o1}I_{h1}}$, the width of the total-reflection domain $2\delta = \overline{I_{o1}I_{o2}}/k = \overline{I'I''}/k$ and the deviation parameter $\eta = -\overline{M_oM_h}/\overline{I_{o1}I_{h1}}$, where M_o and M_h are the intersections with T_o and T_h of the normal to the crystal surface drawn from the extremity of any incident wavevector **OM**.

5.1.4. Standing waves

The various waves in a wavefield are coherent and interfere. In the two-beam case, the intensity of the wavefield, using (5.1.2.14) and (5.1.2.24), is

$$\begin{aligned}|D|^2 &= |D_o|^2 \exp(4\pi\mathbf{K_{oi}}\cdot\mathbf{r}) \\ &\quad \times [1 + |\xi|^2 + 2C|\xi|\cos 2\pi(\mathbf{h}\cdot\mathbf{r} + \Psi)],\end{aligned} \tag{5.1.4.1}$$

where Ψ is the phase of ξ,

$$\xi = |\xi|\exp(i\Psi). \tag{5.1.4.2}$$

Equation (5.1.4.1) shows that the interference between the two waves is the origin of *standing waves*. The corresponding nodes lie on planes such that $\mathbf{h}\cdot\mathbf{r}$ is a constant. These planes are therefore parallel to the diffraction planes and their periodicity is equal to d_{hkl} (defined in the caption for Fig. 5.1.2.1*a*). Their position within the unit cell is given by the value of the phase Ψ.

In the Laue case, Ψ is equal to $\pi + \varphi_h$ for branch 1 and to φ_h for branch 2, where φ_h is the phase of the structure factor, (5.1.2.6). This means that the *nodes* of standing waves lie on the maxima of the *hkl* Fourier component of the electron density for branch 1 while the *anti-nodes* lie on the maxima for branch 2.

In the Bragg case, Ψ varies continuously from $\pi + \varphi_h$ to φ_h as the angle of incidence is varied from the low-angle side to the high-angle side of the reflection domain by rocking the crystal. The nodes lie on the maxima of the *hkl* Fourier components of the electron density on the low-angle side of the rocking curve. As the crystal is rocked, they are progressively shifted by half a lattice spacing until the anti-nodes lie on the maxima of the electron density on the high-angle side of the rocking curve.

Standing waves are the origin of the phenomenon of anomalous absorption, which is one of the specific properties of wavefields (Section 5.1.5). Anomalous scattering is also used for the location of atoms in the unit cell at the vicinity of the crystal surface: when X-rays are absorbed, fluorescent radiation and photoelectrons are emitted. Detection of this emission for a known angular position of the crystal with respect to the rocking curve and therefore for a known value of the phase Ψ enables the emitting atom within the unit cell to be located. The principle of this method is due to Batterman (1964, 1969). For reviews, see Golovchenko *et al.* (1982), Materlik & Zegenhagen (1984), Kovalchuk & Kohn (1986), Bedzyk (1988), Authier (1989), Zegenhagen (1993), Patel (1996), Patel & Fontes (1996), Lagomarsino (1996), Zegenhagen *et al.* (1997), Vartanyants & Zegenhagen (1999) and Vartanyants & Koval'chuk (2001).

5.1.5. Anomalous absorption

It was shown in Section 5.1.2.2 that the wavevectors of a given wavefield all have the same imaginary part (5.1.2.17) and therefore the same absorption coefficient μ (5.1.2.19). Borrmann (1950, 1954) showed that this coefficient is much smaller than the normal one (μ_o) for wavefields whose tie points lie on branch 1 of the dispersion surface and much larger for wavefields whose tie points lie on branch 2. The former case corresponds to the *anomalous transmission effect*, or *Borrmann effect*. As in favourable cases the minimum absorption coefficient may be as low as a few per cent of μ_o, this effect is very important from both a fundamental and a practical point of view.

The physical interpretation of the Borrmann effect is to be found in the standing waves described in Section 5.1.4. When the

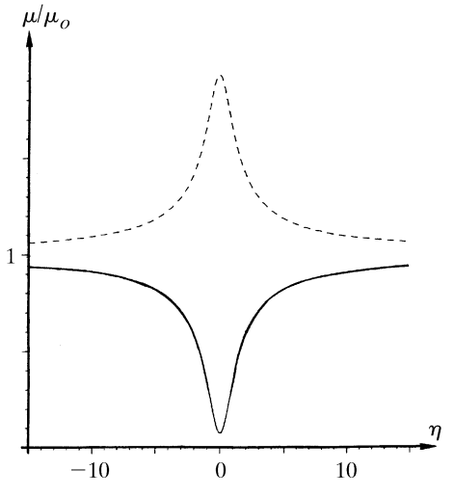


Fig. 5.1.6.1. Variation of the effective absorption with the deviation parameter in the transmission case for the 400 reflection of GaAs using Cu $K\alpha$ radiation. Solid curve: branch 1; broken curve: branch 2.

nodes of the electric field lie on the planes corresponding to the maxima of the *hkl* component of the electron density, the wavefields are absorbed anomalously less than when there is no diffraction. Just the opposite occurs for branch 2 wavefields, whose anti-nodes lie on the maxima of the electron density and which are absorbed more than normal.

The effective absorption coefficient μ is related to the imaginary part of the wavevectors through (5.1.2.19),

$$\mu = -4\pi\gamma_o K_{oi},$$

and to the imaginary part of the ratio of the amplitude of the reflected to the incident wave through

$$\mu = \mu_o - 4\pi X_{oi}, \tag{5.1.5.1}$$

where X_{oi} is the imaginary part of X_o, which, using (5.1.2.24) and (5.1.3.10), is given by

$$X_o = R\lambda|C|\mathrm{S}(\gamma_h)(F_h F_{\bar{h}})^{1/2} \times \{\eta \pm [\eta^2 + \mathrm{S}(\gamma_h)]^{1/2}\}\big/\big[2\pi V(|\gamma|)^{1/2}\big]. \tag{5.1.5.2}$$

Taking the upper sign (+) for the $\pm$ term corresponds to tie points on branch 1 and taking the lower sign (−) corresponds to tie points on branch 2.

The calculation of the imaginary part X_{oi} is different in the Laue and in the Bragg cases. In the former case, the imaginary part of $(\eta^2+1)^{1/2}$ is small and can be approximated while in the latter, the imaginary part of $(\eta^2-1)^{1/2}$ is large when the real part of the deviation parameter, η_r, lies between 1 and −1, and cannot be calculated using the same approximation.

5.1.6. Intensities of plane waves in transmission geometry

5.1.6.1. Absorption coefficient

In transmission geometry, the imaginary part of X_o is small and, using a first-order approximation for the expansion of $(\eta^2+1)^{1/2}$, (5.1.5.1) and (5.1.5.2), the effective absorption coefficient in the absorption case is

$$\mu_j = \mu_o\left[\tfrac{1}{2}(1+\gamma^{-1}) \mp \frac{(\eta_r/2)(1-\gamma^{-1}) + |C|(\gamma^{-1})^{1/2}|F_{ih}/F_{io}|\cos\varphi}{(\eta_r^2+1)^{1/2}}\right], \tag{5.1.6.1}$$

where $\varphi = \varphi_{rh} - \varphi_{ih}$ is the phase difference between F_{rh} and F_{ih} [equation (5.1.2.10)], the upper sign (−) for the $\mp$ term corresponds to branch 1 and the lower sign (+) corresponds to branch 2 of the dispersion surface. In the symmetric Laue case ($\gamma = 1$, reflecting planes normal to the crystal surface), equation (5.1.6.1) reduces to

$$\mu_j = \mu_o\left[1 \mp \frac{|C||F_{ih}/F_{io}|\cos\varphi}{(\eta_r^2+1)^{1/2}}\right].$$

Fig. 5.1.6.1 shows the variations of the effective absorption coefficient μ_j with η_r for wavefields belonging to branches 1 and 2 in the case of the 400 reflection of GaAs with Cu $K\alpha$ radiation. It can be seen that for $\eta_r = 0$ the absorption coefficient for branch 1 becomes significantly smaller than the normal absorption coefficient, μ_o. The minimum absorption coefficient, $\mu_o(1 - |CF_{ih}/F_{io}|\cos\varphi)$, depends on the nature of the reflection through the structure factor and on the temperature through the Debye–Waller factor included in F_{ih} [equation (5.1.2.10*b*)] (Ohtsuki, 1964, 1965). For instance, in diamond-type structures, it is smaller for reflections with even indices than for reflections with odd indices. The influence of temperature is very important when $|F_{ih}/F_{io}|$ is close to one; for example, for germanium 220 and Mo $K\alpha$ radiation, the minimum absorption coefficient at 5 K is reduced to about 1% of its normal value, μ_o (Ludewig, 1969).

5.1.6.2. Boundary conditions for the amplitudes at the entrance surface – intensities of the reflected and refracted waves

Let us consider an infinite plane wave incident on a crystal plane surface of infinite lateral extension. As has been shown in Section 5.1.3, two wavefields are excited in the crystal, with tie points P_1 and P_2, and amplitudes D_{o1}, D_{h1} and D_{o2}, D_{h2}, respectively. Maxwell's boundary conditions (see Section A5.1.1.2 of the Appendix) imply continuity of the tangential component of the electric field and of the normal component of the electric displacement across the boundary. Because the index of refraction is so close to unity, one can assume to a very good approximation that there is continuity of the three components of both the electric field and the electric displacement. As a consequence, it can easily be shown that, *along the entrance surface*, for all components of the electric displacement

$$\begin{aligned} D_o^{(a)} &= D_{o1} + D_{o2} \\ 0 &= D_{h1} + D_{h2}, \end{aligned} \tag{5.1.6.2}$$

where $D_o^{(a)}$ is the amplitude of the incident wave.

Using (5.1.3.11), (5.1.5.2) and (5.1.6.2), it can be shown that the *intensities* of the four waves are

$$\begin{aligned} |D_{oj}|^2 &= |D_o^{(a)}|^2 \exp(-\mu_j z/\gamma_o)\big[(1+\eta_r^2)^{1/2} \mp \eta_r\big]^2 \times [4(1+\eta_r^2)]^{-1}, \\ |D_{hj}|^2 &= |D_o^{(a)}|^2 \exp(-\mu_j z/\gamma_o)|F_h/F_{\bar{h}}|[4\gamma(1+\eta_r^2)]^{-1}; \end{aligned} \tag{5.1.6.3}$$

top sign: $j = 1$; bottom sign: $j = 2$.

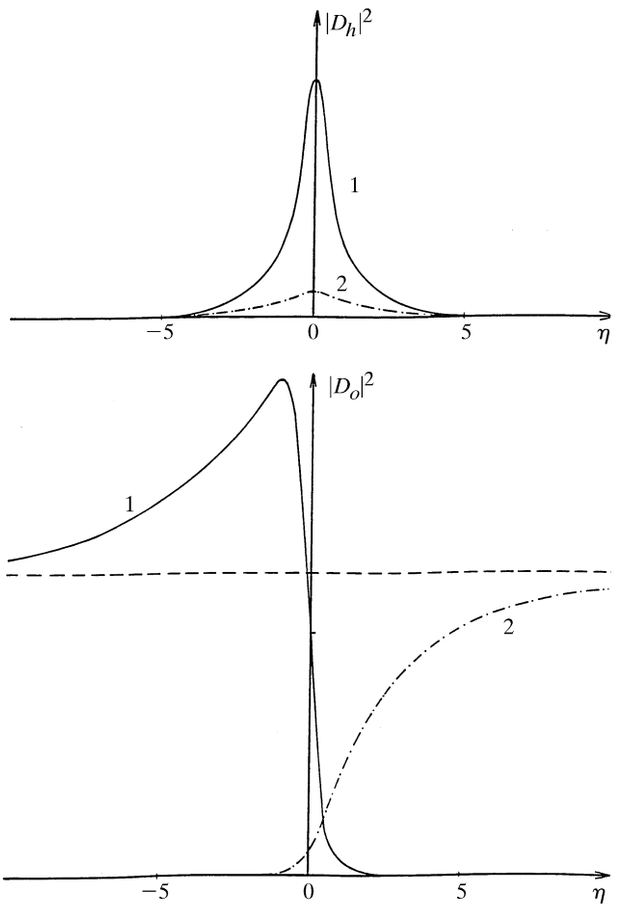


Fig. 5.1.6.2. Variation of the intensities of the reflected and refracted waves in an absorbing crystal for the 220 reflection of Si using Mo $K\alpha$ radiation, t = 1 mm (μt = 1.42). Solid curve: branch 1; dashed curve: branch 2.

Fig. 5.1.6.2 represents the variations of these four intensities with the deviation parameter. Far from the reflection domain, $|D_{h1}|^2$ and $|D_{h2}|^2$ tend toward zero, as is normal, while

$$|D_{o1}|^2 \gg |D_{o2}|^2 \text{ for } \eta_r \Rightarrow -\infty,$$
$$|D_{o1}|^2 \ll |D_{o2}|^2 \text{ for } \eta_r \Rightarrow +\infty.$$

This result shows that the wavefield of highest intensity 'jumps' from one branch of the dispersion surface to the other across the reflection domain. This is an important property of dynamical theory which also holds in the Bragg case and when a wavefield crosses a highly distorted region in a deformed crystal [the so-called *interbranch scattering*: see, for instance, Authier & Balibar (1970) and Authier & Malgrange (1998)].

5.1.6.3. Boundary conditions at the exit surface

5.1.6.3.1. Wavevectors

When a wavefield reaches the exit surface, it breaks up into its two constituent waves. Their wavevectors are obtained by applying again the condition of the continuity of their tangential components along the crystal surface. The extremities, M_j and N_j, of these wavevectors

$$\mathbf{OM}_j = \mathbf{K}_{\mathbf{o}j}^{(d)} \qquad \mathbf{HN}_j = \mathbf{K}_{\mathbf{h}j}^{(d)}$$

lie at the intersections of the spheres of radius k centred at O and H, respectively, with the normal $\mathbf{n}'$ to the crystal exit surface drawn from P_j (j = 1 and 2) (Fig. 5.1.6.3).

If the crystal is wedge-shaped and the normals $\mathbf{n}$ and $\mathbf{n}'$ to the entrance and exit surfaces are not parallel, the wavevectors of the waves generated by the two wavefields are not parallel. This effect is due to the refraction properties associated with the dispersion surface.

5.1.6.3.2. Amplitudes – Pendellösung

We shall assume from now on that the crystal is plane parallel. Two wavefields arrive at any point of the exit surface. Their constituent waves interfere and generate emerging waves in the refracted and reflected directions (Fig. 5.1.6.4). Their respective amplitudes are given by the boundary conditions

$$\begin{aligned} D_o^{(d)} \exp(-2\pi i \mathbf{K}_{\mathbf{o}}^{(d)} \cdot \mathbf{r}) &= D_{o1} \exp(-2\pi i \mathbf{K}_{\mathbf{o}1} \cdot \mathbf{r}) \\ &\quad + D_{o2} \exp(-2\pi i \mathbf{K}_{\mathbf{o}2} \cdot \mathbf{r}) \\ D_h^{(d)} \exp(-2\pi i \mathbf{K}_{\mathbf{h}}^{(d)} \cdot \mathbf{r}) &= D_{h1} \exp(-2\pi i \mathbf{K}_{\mathbf{h}1} \cdot \mathbf{r}) \\ &\quad + D_{h2} \exp(-2\pi i \mathbf{K}_{\mathbf{h}2} \cdot \mathbf{r}), \end{aligned} \tag{5.1.6.4}$$

where $\mathbf{r}$ is the position vector of a point on the exit surface, the origin of phases being taken at the entrance surface.

In a plane-parallel crystal, (5.1.6.4) reduces to

$$D_o^{(d)} = D_{o1} \exp\left(-2\pi i \overline{MP_1} \cdot t\right) + D_{o2} \exp\left(-2\pi i \overline{MP_2} \cdot t\right)$$
$$D_h^{(d)} = D_{h1} \exp\left(-2\pi i \overline{MP_1} \cdot t\right) + D_{h2} \exp\left(-2\pi i \overline{MP_2} \cdot t\right),$$

where t is the crystal thickness.

In a *non-absorbing* crystal, the amplitudes squared are of the form

$$\left|D_o^{(d)}\right|^2 = |D_{o1}|^2 + |D_{o2}|^2 + 2D_{o1}D_{o2} \cos 2\pi \overline{P_2P_1} t.$$

This expression shows that the intensities of the refracted and reflected beams are oscillating functions of crystal thickness. The period of the oscillations is called the *Pendellösung* distance and is

$$\Lambda = 1/\overline{P_2P_1} = \Lambda_L/(1 + \eta_r^2)^{1/2}.$$

5.1.6.4. Reflecting power

For an *absorbing* crystal, the intensities of the reflected and refracted waves are

$$\begin{aligned} \left|D_o^{(d)}\right|^2 &= \left|D_o^{(a)}\right|^2 A_\eta \Big\{ \cosh(2v + \mu_a t) \\ &\quad + \cos\left[2\pi t \Lambda^{-1} - 2\eta_i (1 + \eta_r^2)^{-1/2}\right] \Big\} \\ \left|D_h^{(d)}\right|^2 &= \left|D_o^{(a)}\right|^2 |F_h/F_{\bar{h}}| \gamma^{-1} A_\eta \left[\cosh(\mu_a t) - \cos(2\pi t \Lambda^{-1})\right], \end{aligned} \tag{5.1.6.5}$$

where

$$A_\eta = \left[\exp -\mu_o t(\gamma_o^{-1} + \gamma_h^{-1})\right]/2(1 + \eta_r^2),$$
$$\mu_a = \mu_j \Big[1/2(\gamma_o^{-1} - \gamma_h^{-1})\eta_r + |C||F_{ih}/F_{io}| \cos\varphi/(\gamma_o\gamma_h)^{1/2}\Big](1 + \eta_r^2)^{-1/2},$$
$$v = \operatorname{arg\,sinh} \eta_r$$

and μ_j is given by equation (5.1.6.1).

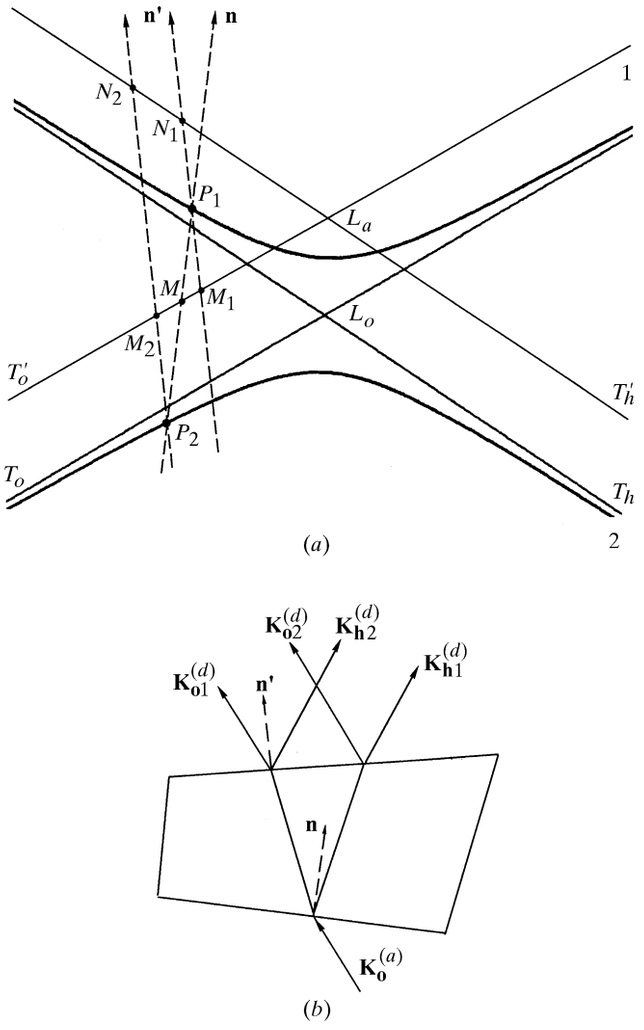


Fig. 5.1.6.3. Boundary condition for the wavevectors at the exit surface. (*a*) Reciprocal space. The wavevectors of the emerging waves are determined by the intersections M_1, M_2, N_1 and N_2 of the normals $\mathbf{n}'$ to the exit surface, drawn from the tie points P_1 and P_2 of the wavefields, with the tangents T'_o and T'_h to the spheres centred at O and H and of radius k, respectively. (*b*) Direct space.

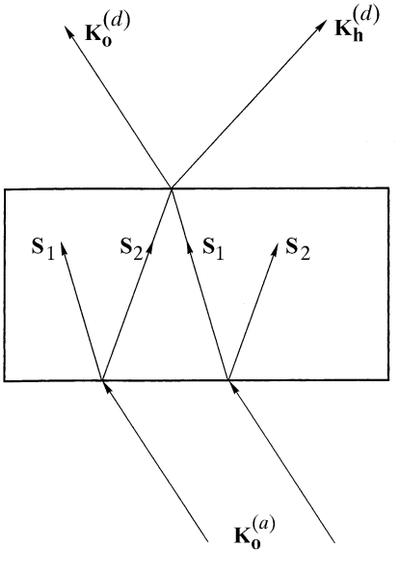


Fig. 5.1.6.4. Decomposition of a wavefield into its two components when it reaches the exit surface. $\mathbf{S}_1$ and $\mathbf{S}_2$ are the Poynting vectors of the two wavefields propagating in the crystal belonging to branches 1 and 2 of the dispersion surface, respectively, and interfering at the exit surface.

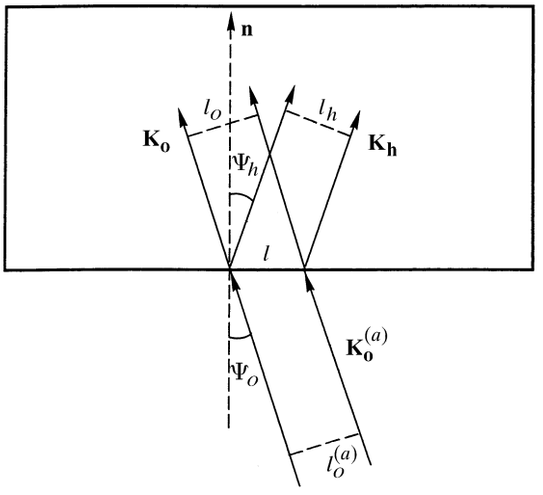


Fig. 5.1.6.5. Cross sections of the incident, $\mathbf{K}_o^{(a)}$, refracted, $\mathbf{K}_o$, and reflected, $\mathbf{K}_h$, waves.

Depending on the absorption coefficient, the cosine terms are more or less important relative to the hyperbolic cosine term and the oscillations due to *Pendellösung* have more or less contrast.

For a non-absorbing crystal, these expressions reduce to

$$\begin{aligned}\left|D_o^{(d)}\right|^2 &= \left|D_o^{(a)}\right|^2\left[\frac{1+2\eta^2+\cos(2\pi t\Lambda^{-1})}{2(1+\eta_r^2)}\right],\\ \left|D_h^{(d)}\right|^2 &= \left|D_o^{(a)}\right|^2\left[\frac{1-\cos(2\pi t\Lambda^{-1})}{2\gamma(1+\eta_r^2)}\right].\end{aligned} \tag{5.1.6.6}$$

What is actually measured in a counter receiving the reflected or the refracted beam is the *reflecting power*, namely the ratio of the energy of the reflected or refracted beam on the one hand and the energy of the incident beam on the other. The energy of a beam is obtained by multiplying its intensity by its cross section. If l is the width of the trace of the beam on the crystal surface, the cross sections of the incident (or refracted) and reflected beams are proportional to (Fig. 5.1.6.5) $l_o = l\gamma_o$ and $l_h = l\gamma_h$, respectively.

The reflecting powers are therefore:

$$\begin{aligned}\text{Refracted beam: } I_o &= l_o\left|D_o^{(d)}\right|^2 \Big/ l_o\left|D_o^{(a)}\right|^2 = \left|D_o^{(d)}\right|^2 \Big/ \left|D_o^{(a)}\right|^2,\\ \text{Reflected beam: } I_h &= l_h\left|D_h^{(d)}\right|^2 \Big/ l_o\left|D_o^{(a)}\right|^2 = \gamma\left|D_h^{(d)}\right|^2 \Big/ \left|D_o^{(a)}\right|^2.\end{aligned} \tag{5.1.6.7}$$

Using (5.1.6.6), it is easy to check that $I_o + I_h = 1$ in the non-absorbing case; that is, that conservation of energy is satisfied. Equations (5.1.6.6) show that there is a periodic exchange of energy between the refracted and the reflected waves as the beam penetrates the crystal; this is why Ewald introduced the expression *Pendellösung*.

The oscillations in the rocking curve were first observed by Lefeld-Sosnowska & Malgrange (1968, 1969). Their periodicity can be used for accurate measurements of the form factor [see, for instance, Bonse & Teworte (1980)]. Fig. 5.1.6.6 shows the shape of the rocking curve for various values of t/Λ_L.

The *width at half-height of the rocking curve*, averaged over the *Pendellösung* oscillations, corresponds in the non-absorbing case to $\Delta\eta = 2$, that is, to $\Delta\theta = 2\delta$, where δ is given by (5.1.3.6).

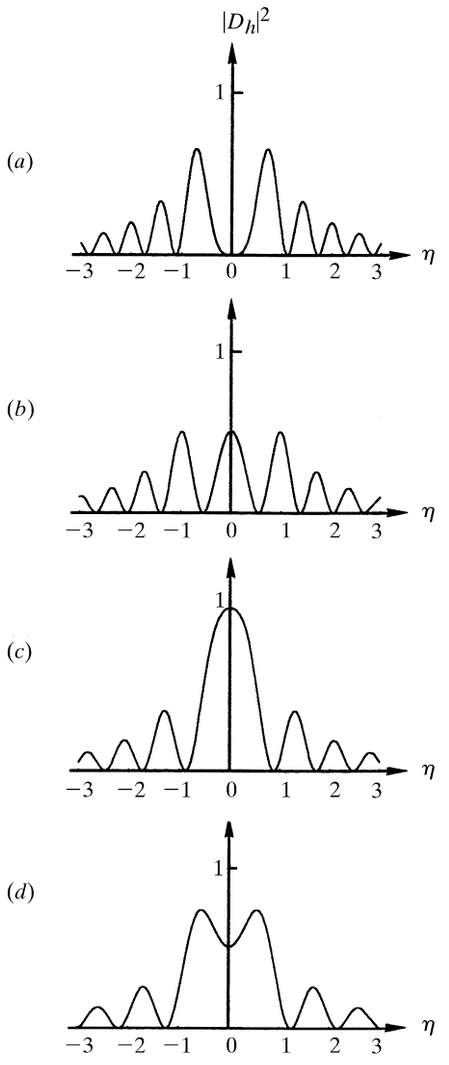


Fig. 5.1.6.6. Theoretical rocking curves in the transmission case for non-absorbing crystals and for various values of t/Λ_L: (*a*) $t/\Lambda_L = 1.25$; (*b*) $t/\Lambda_L = 1.5$; (*c*) $t/\Lambda_L = 1.75$; (*d*) $t/\Lambda_L = 2.0$.

5.1.6.5. Integrated intensity

5.1.6.5.1. Non-absorbing crystals

The integrated intensity is the ratio of the total energy recorded in the counter when the crystal is rocked to the intensity of the incident beam. It is proportional to the area under the line profile:

$$I_{hi} = \int_{-\infty}^{+\infty} I_h \, \mathrm{d}(\Delta\theta). \qquad (5.1.6.8)$$

The integration was performed by von Laue (1960). Using (5.1.3.5), (5.1.6.6) and (5.1.6.7) gives

$$I_{hi} = A \int_0^{2\pi t\Lambda_L^{-1}} J_0(z) \, \mathrm{d}z,$$

where $J_0(z)$ is the zeroth-order Bessel function and

$$A = \frac{R\lambda^2 |CF_h|(\gamma)^{1/2}}{2V \sin 2\theta}.$$

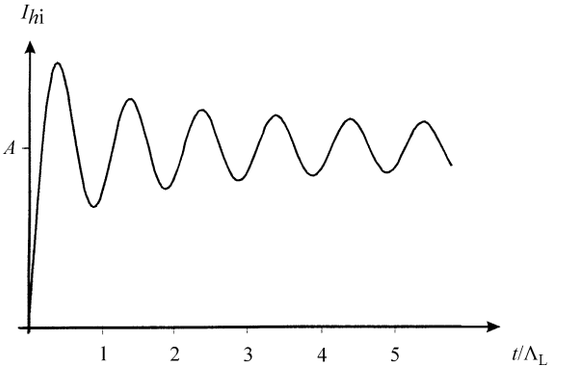


Fig. 5.1.6.7. Variations with crystal thickness of the integrated intensity in the transmission case (no absorption) (arbitrary units). The expression for A is given in the text.

Fig. 5.1.6.7 shows the variations of the integrated intensity with t/Λ_L.

5.1.6.5.2. Absorbing crystals

The integration was performed for absorbing crystals by Kato (1955). The integrated intensity in this case is given by

$$I_{hi} = A|F_h/F_{\bar{h}}| \exp\left[-1/2\mu_o t(\gamma_o^{-1} + \gamma_h^{-1})\right] \times \left[\int_0^{2\pi t\Lambda_L^{-1}} J_0(z) \, \mathrm{d}z - 1 + I_0(\zeta)\right],$$

where

$$\zeta = \mu_o t\left\{\left[|C|^2|F_{ih}/F_{io}|^2 \cos^2\varphi + (\gamma_h - \gamma_o)/(4\gamma_o\gamma_h)\right]/(\gamma_o\gamma_h)\right\}^{1/2}$$

and $I_0(\zeta)$ is a modified Bessel function of zeroth order.

5.1.6.6. Thin crystals – comparison with geometrical theory

Using (5.1.6.6) and (5.1.6.7), the reflecting power of the reflected beam may also be written

$$I_h = \pi^2 t^2 \Lambda_o^{-2} f(\eta),$$

where

$$f(\eta) = \left[\frac{\sin U(1+\eta^2)^{1/2}}{U(1+\eta^2)^{1/2}}\right]^2$$

and

$$U = \pi t \Lambda_o^{-1}.$$

When $t\Lambda_o^{-1}$ is very small, $f(\eta)$ tends asymptotically towards the function

$$f_1(\eta) = \left[\frac{\sin U\eta}{U\eta}\right]^2$$

and I_h towards the value given by geometrical theory. The condition for geometrical theory to apply is, therefore, that the crystal thickness be much smaller than the *Pendellösung* distance. In practice, the two theories agree to within a few per cent for a

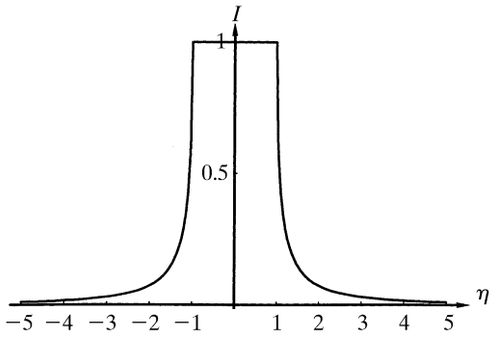


Fig. 5.1.7.1. Theoretical rocking curve in the reflection case for a non-absorbing thick crystal in terms of the deviation parameter.

crystal thickness smaller than or equal to a third of the *Pendellösung* distance [see Authier & Malgrange (1998)].

5.1.7. Intensity of plane waves in reflection geometry

5.1.7.1. Thick crystals

5.1.7.1.1. Non-absorbing crystals

Rocking curve. The geometrical construction in Fig. 5.1.3.5 shows that, in the Bragg case, the normal to the crystal surface drawn from the extremity of the incident wavevector intersects the dispersion surface either at two points of the same branch, P_1, P_1', for branch 1, P_2, P_2' for branch 2, or at imaginary points. It was shown in Section 5.1.2.6 that the propagation of the wavefields inside the crystal is along the normal to the dispersion surface at the corresponding tie points. Fig. 5.1.3.5 shows that this direction is oriented towards the outside of the crystal for tie points P_1' and P_2'. In a very thick crystal, these wavefields cannot exist because there is always a small amount of absorption. One concludes that in the thick-crystal case and in reflection geometry, only one wavefield is excited inside the crystal. It corresponds to branch 1 on the low-angle side of the rocking curve and to branch 2 on the high-angle side. Using the same approximations as in Section 5.1.6.2, the amplitude $\mathbf{D}_\mathbf{h}^{(a)}$ of the wave reflected at the crystal surface is obtained by applying the boundary conditions, which are particularly simple in this case:

$$\mathbf{D}_\mathbf{o} = \mathbf{D}_\mathbf{o}^{(a)},\ \mathbf{D}_\mathbf{h}^{(a)} = \mathbf{D}_\mathbf{h}.$$

The reflecting power is given by an expression similar to (5.1.6.7):

$$I_h = |\gamma||\xi_j|^2,$$

where the expression for ξ_j is given by (5.1.3.12), and $j = 1$ or 2 depending on which wavefield propagates towards the inside of the crystal. When the normal to the entrance surface intersects the dispersion surface at imaginary points, *i.e.* when $-1 < \eta < +1$,

$$|\xi|^2 = |\gamma|^{-1},\ I_h = 1, \tag{5.1.7.1}$$

and there is *total reflection*. Outside the total-reflection domain, the reflecting power is given by

$$I_h = \left[|\eta| - (\eta^2 - 1)^{1/2}\right]^2. \tag{5.1.7.2}$$

The rocking curve has the well known top-hat shape (Fig. 5.1.7.1). Far from the total-reflection domain, the curve can be approximated by the function

$$I_h \simeq 1/(4\eta^2).$$

Width of the total-reflection domain. The width of the total-reflection domain is equal to $\Delta\eta = 2$ and its angular width is therefore equal, using (5.1.3.5), to 2δ, where δ is given by (5.1.3.6). It is proportional to the structure factor, the polarization factor C and the square root of the asymmetry factor $|\gamma|$. Using an asymmetric reflection, it is therefore possible to decrease the width at wish. This is used in monochromators to produce a pseudo plane wave [see, for instance, Kikuta & Kohra (1970)]. It is possible to deduce the value of the form factor from very accurate measurements of the rocking curve; see, for instance, Kikuta (1971).

Integrated intensity. The integrated intensity is defined by (5.1.6.8):

$$I_{h\mathrm{i}} = 8\delta/3. \tag{5.1.7.3}$$

Penetration depth. Within the domain of total reflection, there are two wavefields propagating inside the crystal with imaginary wavevectors, one towards the inside of the crystal and the other one in the opposite direction, so that they cancel out and, globally, no energy penetrates the crystal. The absorption coefficient of the waves penetrating the crystal is

$$\mu = -4\pi K_{o\mathrm{i}}\gamma_o = 2\pi\gamma_o(1 - \eta^2)^{1/2}/\Lambda_\mathrm{B}, \tag{5.1.7.4}$$

where Λ_B is the value taken by Λ_o [equation (5.1.3.8)] in the Bragg case.

The penetration depth is a minimum at the middle of the reflection domain and at this point it is equal to $\Lambda_\mathrm{B}/2\pi$. This attenuation effect is called *extinction*, and Λ_B is called the *extinction length.* It is a specific property owing to the existence of wavefields. The resulting propagation direction of energy is parallel to the crystal surface, but with a cross section equal to zero: it is an *evanescent wave* [see, for instance, Cowan *et al.* (1986)].

5.1.7.1.2. Absorbing crystals

Rocking curve. Since the sign of γ is negative, $[\eta^2 + \mathrm{S}(\gamma h)]^{1/2}$ in (5.1.3.10) has a very large imaginary part when $|\eta_\mathrm{r}| \leq 1$. It cannot be calculated using the same approximations as in the Laue case. Let us set

$$Z\exp(i\Psi') = \eta \mp (\eta^2 - 1)^{1/2}. \tag{5.1.7.5}$$

The reflecting power is

$$I_h = (F_h/F_{\bar{\mathbf{h}}})^{1/2}Z^2, \tag{5.1.7.6}$$

where $Z = [L - (L^2 - 1)^{1/2}]^{1/2}$, $L = |\eta|^2 + \rho^2$ and $\rho = |\eta^2 - 1|$ is the modulus of expression (5.1.7.5) where the sign is chosen in such a way that Z is smaller than 1.

The expression for the reflected intensity in the absorbing Bragg case was first given by Prins (1930). The way of representing it given here was first used by Hirsch & Ramachandran (1950). The properties of the rocking curve have been described by Fingerland (1971).

There is no longer a total-reflection domain and energy penetrates the crystal at all incidence angles, although with a very

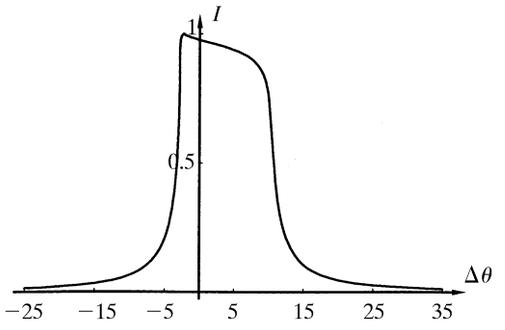


Fig. 5.1.7.2. Theoretical rocking curve in the reflection case for a thick absorbing crystal. The 400 reflection of GaAs using Cu $K\alpha$ radiation is shown.

high absorption coefficient within the domain $|\eta_r| \leq 1$. Fig. 5.1.7.2 gives an example of a rocking curve for a thick absorbing crystal. It was first observed by Renninger (1955). The shape is asymmetric and is due to the anomalous-absorption effect: it is lower than normal on the low-angle side, which is associated with wavefields belonging to branch 1 of the dispersion surface, and larger than normal on the high-angle side, which is associated with branch 2 wavefields. The amount of asymmetry depends on the value of the ratio A/B of the coefficients in the expression for the imaginary part of the deviation parameter in (5.1.3.7): the smaller this ratio, the more important the asymmetry.

Absorption coefficient. The effective absorption coefficient, taking into account both the Borrmann effect and extinction, is given by (Authier, 1986)

$$\mu = \mu_o + 2(|F_h F_{\bar{h}}|)^{1/2} \frac{R\lambda}{V(|\gamma|)^{1/2}} Z \sin(\beta + \Psi'),$$

where β is defined in equation (5.1.3.7) and Ψ' in equation (5.1.7.5), and where the sign is chosen in such a way that Z converges. Fig. 5.1.7.3 shows the variation of the penetration depth $z_o = \gamma_o/\mu$ with the deviation parameter.

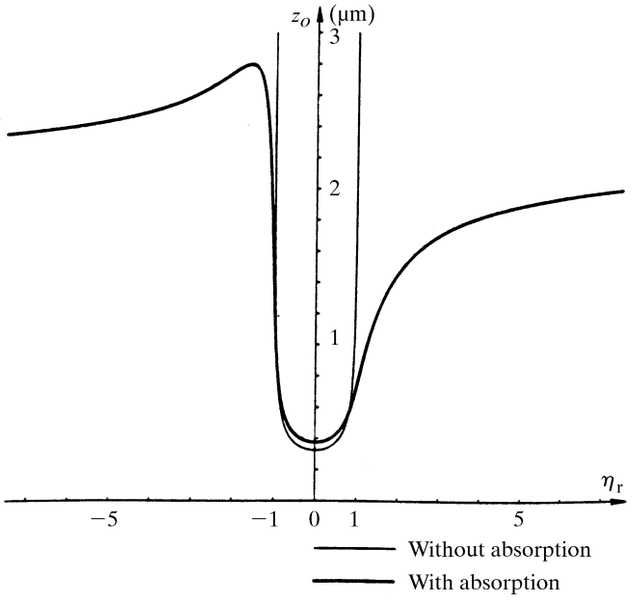


Fig. 5.1.7.3. Bragg case: thick crystals. Variation of the penetration depth with incidence angle (represented here by the dimensionless deviation parameter η). Thin curve: without absorption; thick curve: with absorption for the 400 reflection of GaAs using Cu $K\alpha$ radiation.

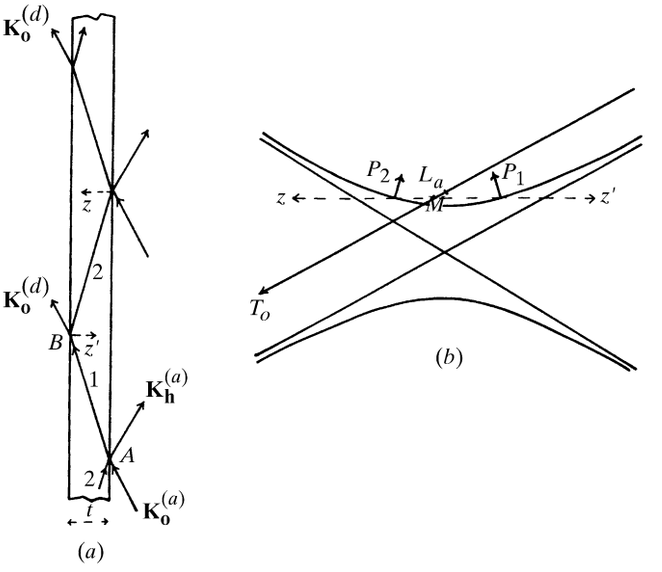


Fig. 5.1.7.4. Bragg case: thin crystals. Two wavefields propagate in the crystal. (*a*) Direct space; (*b*) reciprocal space.

5.1.7.2. Thin crystals

5.1.7.2.1. Non-absorbing crystals

Boundary conditions. If the crystal is thin, the wavefield created at the reflecting surface at A and penetrating inside can reach the back surface at B (Fig. 5.1.7.4*a*). The *incident* direction there points towards the *outside* of the crystal, while the *reflected* direction points towards the inside. The wavefield propagating along AB will therefore generate at B:

(i) *a partially transmitted wave* outside the crystal, $\mathbf{D}_\mathbf{o}^{(d)} \exp(-2\pi i \mathbf{K}_\mathbf{o}^{(d)} \cdot \mathbf{r})$;

(ii) *a partially reflected wavefield* inside the crystal.

The corresponding tiepoints are obtained by applying the usual condition of the continuity of the tangential components of wavevectors (Fig. 5.1.7.4*b*). If the crystal is a plane-parallel slab, these points are M and P_2, respectively, and $\mathbf{K}_\mathbf{o}^{(d)} = \mathbf{K}_\mathbf{o}^{(a)}$.

The boundary conditions are then written:

(i) entrance surface:

$$D_{o1} + D_{o2} = D_o^{(a)}, \;\; D_{h1} + D_{h2} = D_h^{(a)};$$

(ii) back surface:

$$\begin{aligned} &D_{o1} \exp(-2\pi i \mathbf{K}_{\mathbf{o}1} \cdot \mathbf{r}) + D_{o2} \exp(-2\pi i \mathbf{K}_{\mathbf{o}2} \cdot \mathbf{r}) \\ &\quad = D_o^{(a)} \exp\big(-2\pi i \mathbf{K}_\mathbf{o}^{(d)} \cdot \mathbf{r}\big), \\ &D_{h1} \exp(-2\pi i \mathbf{K}_{\mathbf{h}1} \cdot \mathbf{r}) + D_{h2} \exp(-2\pi i \mathbf{K}_{\mathbf{h}2} \cdot \mathbf{r}) = 0. \end{aligned}$$

Rocking curve. Using (5.1.3.10), it can be shown that the expressions for the intensities reflected at the entrance surface and transmitted at the back surface, I_h and I_o, respectively, are given by different expressions within total reflection and outside it:

(i) $|\eta| < 1$:

$$\begin{aligned} I_h &= |\gamma| \frac{|D_h^{(a)}|^2}{|D_o^{(a)}|^2} = \frac{\cosh^2[\pi(t/\Lambda_\mathrm{B})(1-\eta^2)^{1/2}] - 1}{\cosh^2[\pi(t/\Lambda_\mathrm{B})(1-\eta^2)^{1/2}] - \eta^2} \\ I_o &= \frac{|D_o^{(d)}|^2}{|D_o^{(a)}|^2} = \frac{1-\eta^2}{\cosh^2[\pi(t/\Lambda_\mathrm{B})(1-\eta^2)^{1/2}] - \eta^2}, \end{aligned} \qquad (5.1.7.7a)$$

where Λ_B is the value taken by Λ_o [equation (5.1.3.8)] in the Bragg case.

There is no longer a total-reflection domain but the extinction effect still exists, as is shown by the hyperbolic cosine term. The maximum height of the rocking curve decreases as the thickness of the crystal decreases.

(ii) $|\eta| > 1$:

$$I_h = \frac{1 - \cos^2[\pi(t/\Lambda_B)(\eta^2-1)^{1/2}]}{\eta^2 - \cos^2[\pi(t/\Lambda_B)(\eta^2-1)^{1/2}]},$$
$$I_o = \frac{\eta^2 - 1}{\eta^2 - \cos^2[\pi(t/\Lambda_B)(\eta^2-1)^{1/2}]}. \quad (5.1.7.7b)$$

The cosine terms show that the two wavefields propagating within the crystal interfere, giving rise to *Pendellösung* fringes in the rocking curve. These fringes were observed for the first time by Batterman & Hildebrandt (1967, 1968). The angular positions of the minima of the reflected beam are given by

$$\eta = \mp(K^2\Lambda_B^2 t^{-2} + 1)^{1/2},$$

where K is an integer.

Integrated intensity. The integrated intensity is

$$I_{hi} = \pi\delta \tanh[\pi t/\Lambda_B], \quad (5.1.7.8)$$

where t is the crystal thickness. When this thickness becomes very large, the integrated intensity tends towards

$$I_{hi} = \pi\delta. \quad (5.1.7.9)$$

This expression differs from (5.1.7.3) by the factor π, which appears here in place of 8/3. von Laue (1960) pointed out that because of the differences between the two expressions for the reflecting power, (5.1.7.2) and (5.1.7.7*b*), perfect agreement could not be expected. Since some absorption is always present, expression (5.1.7.3), which includes the factor 8/3, should be used for very thick crystals. In the presence of absorption, however, expression (5.1.7.8) for the reflected intensity for thin crystals does tend towards that for thick crystals as the crystal thickness increases.

Comparison with geometrical theory. When t/Λ_B is very small (thin crystals or weak reflections), (5.1.7.8) tends towards

$$I_{hi} = R^2\lambda^2 t|F_h|^2 \big/ (V^2\gamma_o \sin 2\theta), \quad (5.1.7.10)$$

which is the expression given by geometrical theory. If we call this intensity I_{hi}(geom.), comparison of expressions (5.1.7.8) and (5.1.7.10) shows that the integrated intensity for crystals of intermediate thickness can be written

$$I_{hi} = I_{hi}(\text{geom.})\frac{\tanh(\pi t/\Lambda_B)}{(\pi t/\Lambda_B)}, \quad (5.1.7.11)$$

which is the expression given by Darwin (1922) for primary extinction.

5.1.7.2.2. Absorbing crystals

Reflected intensity. The intensity of the reflected wave for a thin absorbing crystal is

$$I_h = |\gamma|\left|\frac{D_h^{(a)}}{D_o^{(a)}}\right|^2 = \left|\frac{F_h}{F_{\bar{h}}}\right| \frac{\cosh 2b - \cos 2a}{L\cosh 2b + (L^2-1)^{1/2}\sinh 2b - \cos(2a + 2\Psi')}, \quad (5.1.7.12)$$

where

$$2a = [\pi t/\Lambda_B]\rho\cos(\beta+\omega),$$
$$2b = [\pi t/\Lambda_B]\rho\sin(\beta+\omega).$$

L, ρ and ψ' are defined in (5.1.7.5), β is defined in (5.1.3.7) and ω is the phase angle of $(\eta^2-1)^{1/2}$.

Comparison with geometrical theory. When t/Λ_B decreases towards zero, expression (5.1.7.12) tends towards $[\sin(\pi t\eta/\Lambda_B)/\eta]^2$; using (5.1.3.5) and (5.1.3.8), it can be shown that expression (5.1.7.12) can be written, in the non-absorbing symmetric case, as

$$I_h = \frac{R^2\lambda^2 C^2|F_h|^2 t^2}{V^2\sin^2\theta}\left\{\frac{\sin[2\pi k\cos(\theta)t\Delta\theta]}{[2\pi k\cos(\theta)t\Delta\theta]}\right\}^2, \quad (5.1.7.13)$$

where d is the lattice spacing and $\Delta\theta$ is the difference between the angle of incidence and the middle of the reflection domain. This expression is the classical expression given by geometrical theory [see, for instance, James (1950)].

5.1.8. Real waves

5.1.8.1. Introduction

The preceding sections have dealt with the diffraction of a plane wave by a semi-infinite perfect crystal. This situation is actually never encountered in practice, although with various devices, in particular using synchrotron radiation, it is possible to produce highly collimated monochromated waves which behave like pseudo plane waves. The wave from an X-ray tube is best represented by a spherical wave. The first experimental proof of this fact is due to Kato & Lang (1959) in the transmission case. Kato extended the dynamical theory to spherical waves for non-absorbing (Kato, 1961*a*,*b*) and absorbing crystals (Kato, 1968*a*,*b*). He expanded the incident spherical wave into plane waves by a Fourier transform, applied plane-wave dynamical theory to each of these components and took the Fourier transform of the result again in order to obtain the final solution. Another method for treating the problem was used by Takagi (1962, 1969), who solved the propagation equation in a medium where the lateral extension of the incident wave is limited and where the wave amplitudes depend on the lateral coordinates. He showed that in this case the set of fundamental linear equations (5.1.2.20) should be replaced by a set of partial differential equations. This treatment can be applied equally well to a perfect or to an imperfect crystal. In the case of a perfect crystal, Takagi showed that these equations have an analytical solution that is identical with Kato's result. Uragami (1969, 1970) observed the spherical wave in the Bragg (reflection) case, interpreting the observed intensity distribution using Takagi's theory. Saka *et al.* (1973) subsequently extended Kato's theory to the Bragg case.

Without using any mathematical treatment, it is possible to make some elementary remarks by considering the fact that the divergence of the incident beam falling on the crystal from the source is much larger than the angular width of the reflection domain. Fig. 5.1.8.1(*a*) shows a spatially collimated beam falling

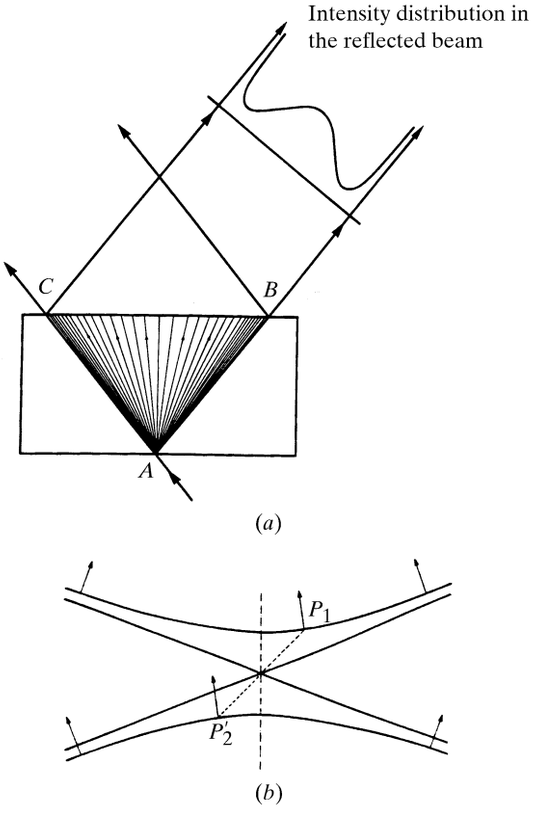


Fig. 5.1.8.1. Borrmann triangle. When the incident beam is divergent, the whole dispersion surface is excited and the wavefields excited inside the crystal propagate within a triangle filling all the space between the incident direction, AC, and the reflected direction, AB. Along any direction Ap within this triangle two wavefields propagate, having as tie points two conjugate points, P and P', at the extremities of a diameter of the dispersion surface. (*a*) Direct space; (*b*) reciprocal space.

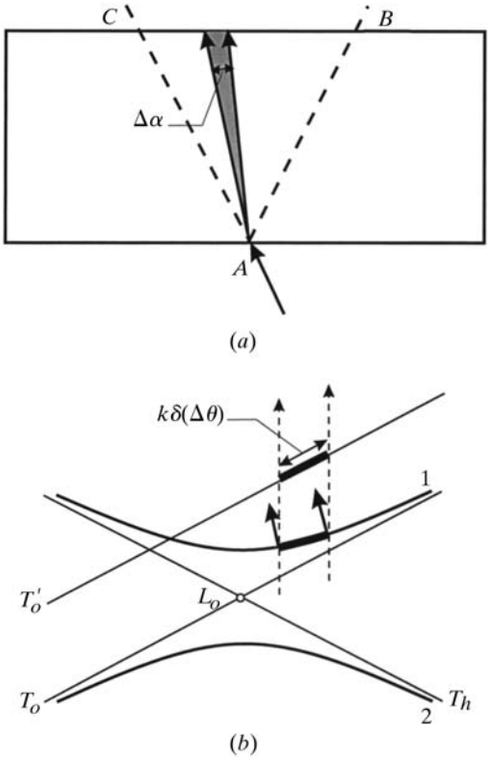


Fig. 5.1.8.2. Packet of wavefields of divergence $\Delta\alpha$ excited in the crystal by an incident wavepacket of angular width $\delta(\Delta\theta)$. (*a*) Direct space; (*b*) reciprocal space.

on a crystal in the transmission case and Fig. 5.1.8.1(*b*) represents the corresponding situation in reciprocal space. Since the divergence of the incident beam is larger than the angular width of the dispersion surface, the plane waves of its Fourier expansion will excite every point of both branches of the dispersion surface. The propagation directions of the corresponding wavefields will cover the angular range between those of the incident and reflected beams (Fig. 5.1.8.1*a*) and fill what is called the *Borrmann triangle*. The intensity distribution within this triangle has interesting properties, as described in the next two sections.

5.1.8.2. Borrmann triangle

The first property of the Borrmann triangle is that the angular density of the wavefield paths is inversely proportional to the curvature of the dispersion surface around their tie points. Let us consider an incident wavepacket of angular width $\delta(\Delta\theta)$. It will generate a packet of wavefields propagating within the Borrmann triangle. The angular width $\Delta\alpha$ (Fig. 5.1.8.2) between the paths of the corresponding wavefields is related to the radius of curvature $\mathcal{R}$ of the dispersion surface by

$$\mathcal{A} = \Delta\alpha/\delta(\Delta\theta) = k\cos\theta(\mathcal{R}\cos\alpha), \qquad (5.1.8.1)$$

where α is the angle between the wavefield path and the lattice planes [equation (5.1.2.26)] and $\mathcal{A}$ is called the amplification ratio. In the middle of the reflecting domain, the radius of curvature of the dispersion surface is very much shorter than its value, k, far from it (about 10^4 times shorter) and the amplification ratio is therefore very large. As a consequence, the energy of a wavepacket of width $\delta(\Delta\theta)$ in reciprocal space is spread in direct space over an angle $\Delta\alpha$ given by (5.1.8.1). The intensity distribution on the exit surface BC (Fig. 5.1.8.1*a*) is therefore proportional to $I_h/\mathcal{A}$. It is represented in Fig. 5.1.8.3 for several values of the absorption coefficient:

(i) Small values of $\mu_o t$ (less than 2 or 3) (Fig. 5.1.8.3*a*). The intensity distribution presents a wide minimum in the centre where the density of wavefields is small and increases very sharply at the edges where the density of wavefields is large, although it is the reverse for the reflecting power I_{hj}. This effect, called the *margin effect*, was predicted qualitatively by Borrmann (1959) and von Laue (1960), demonstrated experimentally by Kato & Lang (1959), and calculated by Kato (1960).

(ii) Large values of $\mu_o t$ (of the order of 6 or more) (Fig. 5.1.8.3*b*). The predominant factor is now anomalous absorption. The wavefields propagating along the edges of the Borrmann triangle undergo normal absorption, while those propagating parallel to the lattice planes (or nearly parallel) correspond to tie points in the centre of the dispersion surface and undergo anomalously low absorption. The intensity distribution now has a maximum in the centre. For values of μt larger than 10 or so, practically only the wavefields propagating parallel to the lattice planes go through the crystal, which acts as a wave guide: this is the Borrmann effect.

5.1.8.3. Spherical-wave Pendellösung

Fig. 5.1.8.4 shows that along any path Ap inside the Borrmann triangle two wavefields propagate, one with tie point P_1, on branch 1, the other with the point P'_2, on branch 2. These two points lie on the extremities of a diameter of the dispersion surface. The two wavefields interfere, giving rise to *Pendellösung* fringes, which were first observed by Kato & Lang (1959), and calculated by Kato (1961*b*). These fringes are of course quite different from the plane-wave *Pendellösung* fringes predicted by Ewald (Section 5.1.6.3) because the tie points of the interfering wavefields are different and their period is also different, but they have in common the fact that they result from interference

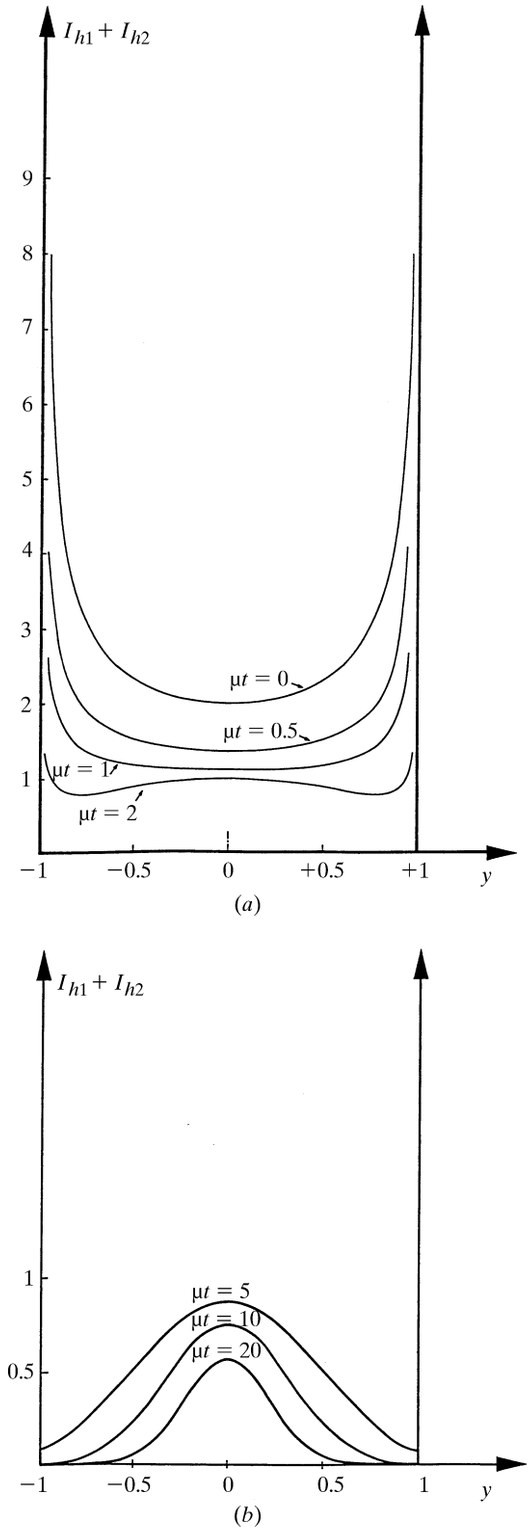


Fig. 5.1.8.3. Intensity distribution along the base of the Borrmann triangle. y is a normalized coordinate along BC. (*a*) Small values of μt. The interference (spherical-wave *Pendellösung*) between branch 1 and branch 2 is neglected. (*b*) Large values of μt.

between wavefields belonging to different branches of the dispersion surface.

Kato has shown that the intensity distribution at any point at the base of the Borrmann triangle is proportional to

$$\left\{J_o\left[A(x_o x_h)^{1/2}\right]\right\}^2,$$

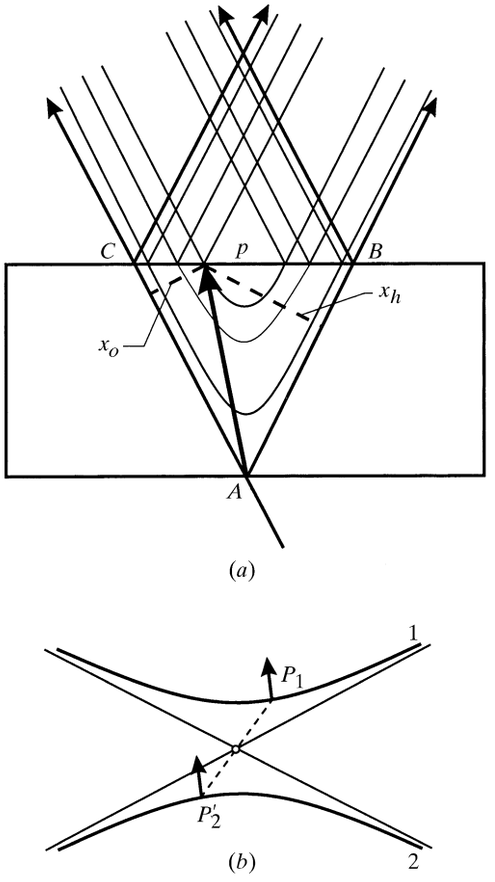


Fig. 5.1.8.4. Interference at the origin of the *Pendellösung* fringes in the case of an incident spherical wave. (*a*) Direct space; (*b*) reciprocal space.

where $A = 2\pi(\gamma_o\gamma_h)^{1/2}/(\Lambda_L \sin\theta)$ and x_o and x_h are the distances of p from the sides AB and AC of the Borrmann triangle (Fig. 5.1.8.4). The equal-intensity fringes are therefore located along the locus of the points in the triangle for which the product of the distances to the sides is constant, that is hyperbolas having AB and AC as asymptotes (Fig. 5.1.8.4*b*). These fringes can be observed on a section topograph of a wedge-shaped crystal (Fig. 5.1.8.5). The technique of section topography is described in *IT* C, Section 2.7.2.2. The *Pendellösung* distance Λ_L depends on the polarization state [see equation (5.1.3.8)]. If the incident wave is unpolarized, one observes the superposition of the *Pendellösung* fringes corresponding to the two states of polarization, parallel and perpendicular to the plane of incidence. This results in a beat effect, which is clearly visible in Fig. 5.1.8.5.

APPENDIX A5.1.1
Basic equations

A5.1.1.1. Dielectric susceptibility – classical derivation

Under the influence of the incident electromagnetic radiation, the medium becomes polarized. The dielectric susceptibility, which relates this polarization to the electric field, thus characterizes the interaction of the medium and the electromagnetic wave. The classical derivation of the dielectric susceptibility, χ, which is summarized here is only valid for a very high frequency which is also far from an absorption edge. Let us consider an electromagnetic wave,

$$\mathbf{E} = \mathbf{E_o} \exp 2\pi i(\nu t - \mathbf{k}\cdot\mathbf{r}),$$

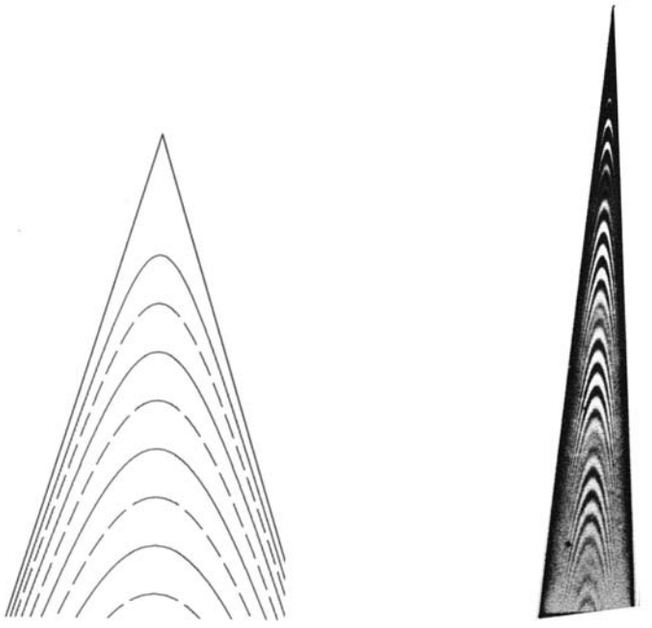


Fig. 5.1.8.5. Spherical-wave *Pendellösung* fringes observed on a wedge-shaped crystal. (*a*) Computer simulation (solid lines: maxima; dashed lines: minima). (*b*) X-ray section topograph of a wedge-shaped silicon crystal (444 reflection, Mo $K\alpha$ radiation).

incident on a bound electron. The electron behaves as if it were held by a spring with a linear restoring force and is an oscillator with a resonant frequency ν_o. The equation of its motion is written in the following way:

$$m\,\mathrm{d}^2\mathbf{a}/\mathrm{d}t^2 = -4\pi^2\nu_o^2 m\mathbf{a} = \mathbf{F},$$

where the driving force $\mathbf{F}$ is due to the electric field of the wave and is equal to $-e\mathbf{E}$. The magnetic interaction is neglected here.

The solution of the equation of motion is

$$\mathbf{a} = -e\mathbf{E}/[4\pi^2 m(\nu_o^2 - \nu^2)].$$

The resonant frequencies of the electrons in atoms are of the order of the ultraviolet frequencies and are therefore much smaller than X-ray frequencies. They can be neglected and the expression of the amplitude of the electron reduces to

$$\mathbf{a} = e\mathbf{E}/(4\pi^2\nu^2).$$

The dipolar moment is therefore

$$\mathcal{M} = -e\mathbf{a} = -e^2\mathbf{E}/(4\pi^2\nu^2 m).$$

von Laue assumes that the negative charge is distributed continuously all over space and that the charge of a volume element $\mathrm{d}\tau$ is $-e\rho\,\mathrm{d}\tau$, where ρ is the electronic density. The electric moment of the volume element is

$$\mathrm{d}\mathcal{M} = -e^2\rho\mathbf{E}\,\mathrm{d}\tau/(4\pi^2\nu^2 m).$$

The polarization is equal to the moment per unit volume:

$$\mathbf{P} = \mathrm{d}\mathcal{M}/\mathrm{d}\tau = -e^2\rho\mathbf{E}/(4\pi^2\nu^2 m).$$

It is related to the electric field and electric displacement through

$$\mathbf{D} = \varepsilon_o\mathbf{E} + \mathbf{P} = \varepsilon_o(1+\chi)\mathbf{E}. \qquad \text{(A5.1.1.1)}$$

We finally obtain the expression of the dielectric susceptibility,

$$\chi = -e^2\rho/(4\pi^2\varepsilon_o\nu^2 m) = -R\lambda^2\rho/\pi, \qquad \text{(A5.1.1.2)}$$

where $R = e^2/(4\pi\varepsilon_o mc^2)$ ($= 2.81794 \times 10^{-15}$ m) is the classical radius of the electron.

A5.1.1.2. Maxwell's equations

The electromagnetic field is represented by two vectors, $\mathbf{E}$ and $\mathbf{B}$, which are the electric field and the magnetic induction, respectively. To describe the interaction of the field with matter, three other vectors must be taken into account, the electrical current density, $\mathbf{j}$, the electric displacement, $\mathbf{D}$, and the magnetic field, $\mathbf{H}$. The space and time derivatives of these vectors are related in a continuous medium by Maxwell's equations:

$$\begin{aligned} \mathrm{curl}\ \mathbf{E} &= -\partial\mathbf{B}/\partial t \\ \mathrm{curl}\ \mathbf{H} &= \partial\mathbf{D}/\partial t + \mathbf{j} \\ \mathrm{div}\ \mathbf{D} &= \rho \\ \mathrm{div}\ \mathbf{B} &= 0, \end{aligned} \qquad \text{(A5.1.1.3)}$$

where ρ is the electric charge density.

The electric field and the electric displacement on the one hand, and the magnetic field and the magnetic induction on the other hand, are related by the so-called *material relations*, which describe the reaction of the medium to the electromagnetic field:

$$\begin{aligned} \mathbf{D} &= \varepsilon\mathbf{E} \\ \mathbf{B} &= \mu\mathbf{H}, \end{aligned}$$

where ε and μ are the dielectric constant and the magnetic permeability, respectively.

These equations are complemented by the following boundary conditions at the surface between two neighbouring media:

$$\begin{aligned} \mathbf{E}_{T1} - \mathbf{E}_{T2} &= 0 \quad \mathbf{D}_{N1} - \mathbf{D}_{N2} = 0 \\ \mathbf{H}_{T1} - \mathbf{H}_{T2} &= 0 \quad \mathbf{B}_{N1} - \mathbf{B}_{N2} = 0. \end{aligned} \qquad \text{(A5.1.1.4)}$$

From the second and the third equations of (A5.1.1.3), and using the identity

$$\mathrm{div}\ (\mathrm{curl}\ \mathbf{y}) = 0,$$

it follows that

$$\mathrm{div}\ \mathbf{j} + \partial\rho/\partial t = 0. \qquad \text{(A5.1.1.5)}$$

A5.1.1.3. Propagation equation

In a vacuum, ρ and $\mathbf{j}$ are equal to zero, and the first two equations of (A5.1.1.3) can be written

$$\text{curl } \mathbf{E} = -\mu_0 \partial \mathbf{E}/\partial t$$
$$\text{curl } \mathbf{H} = \varepsilon_0 \partial \mathbf{H}/\partial t,$$

where ε_0 and μ_0 are the dielectric constant and the magnetic permeability of a vacuum, respectively.

By taking the curl of both sides of the second equation, it follows that

$$\text{curl curl } \mathbf{E} = -\varepsilon_0 \mu_0 \partial^2 \mathbf{E}/\partial t^2.$$

Using the identity curl curl $\mathbf{E}$ = grad div $\mathbf{E} - \Delta \mathbf{E}$, the relation $\varepsilon_0 \mu_0 = 1/c^2$, where c is the velocity of light, and noting that div $\mathbf{E}$ = div $\mathbf{D} = 0$, one finally obtains the equation of propagation of an electromagnetic wave in a vacuum:

$$\Delta \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}. \qquad \text{(A5.1.1.6)}$$

Its simplest solution is a plane wave:

$$\mathbf{E} = \mathbf{E_0} \exp 2\pi i(\nu t - \mathbf{k} \cdot \mathbf{r}),$$

of which the wavenumber $k = 1/\lambda$ and the frequency ν are related by

$$k = \nu/c.$$

The basic properties of the electromagnetic field are described, for instance, in Born & Wolf (1983).

The propagation equation of X-rays in a crystalline medium is derived following von Laue (1960). The interaction of X-rays with charged particles is inversely proportional to the mass of the particle and the interaction with the nuclei can be neglected. As a first approximation, it is also assumed that the magnetic interaction of X-rays with matter is neglected, and that the magnetic permeability μ can be taken as equal to the magnetic permeability of a vacuum, μ_0. It is further assumed that the negative and positive charges are both continuously distributed and compensate each other in such a way that there is neutrality and no current everywhere: ρ and $\mathbf{j}$ are equal to zero and div $\mathbf{D}$ is therefore also equal to zero. The electric displacement is related to the electric field by (A5.1.1.1) and the electric part of the interaction of X-rays with matter is expressed through the dielectric susceptibility χ, which is given by (A5.1.1.2). This quantity is proportional to the electron density and varies with the space coordinates. It is therefore concluded that div $\mathbf{E}$ is different from zero, as opposed to what happens in a vacuum. For this reason, the propagation equation of X-rays in a crystalline medium is expressed in terms of the electric displacement rather than in terms of the electric field. It is obtained by eliminating $\mathbf{H}$, $\mathbf{B}$ and $\mathbf{E}$ in Maxwell's equations and taking into account the above assumptions:

$$\Delta \mathbf{D} + \text{curl curl } \chi \mathbf{D} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}. \qquad \text{(A5.1.1.7)}$$

Only coherent scattering is taken into account here, that is, scattering without frequency change. The solution is therefore a wave of the form

$$\mathbf{D}(\mathbf{r}) \exp 2\pi i \nu t.$$

By replacing $\mathbf{D}$ with this expression in equation (A5.1.1.7), one finally obtains the propagation equation (5.1.2.2) (Section 5.1.2.1).

In a crystalline medium, χ is a triply periodic function of the space coordinates and the solutions of this equation are given in terms of Fourier series which can be interpreted as sums of electromagnetic plane waves. Each of these waves is characterized by its wavevector, $\mathbf{K_h}$, its electric displacement, $\mathbf{D_h}$, its electric field, $\mathbf{E_h}$, and its magnetic field, $\mathbf{H_h}$. It can be shown that, as a consequence of the fact that div $\mathbf{D} = 0$ and div $\mathbf{E} \neq 0$, $\mathbf{D_h}$ is a transverse wave ($\mathbf{D_h}$, $\mathbf{H_h}$ and $\mathbf{K_h}$ and are mutually orthogonal) while $\mathbf{E_h}$ is not. The electric displacement is therefore a more suitable vector for describing the state of the field inside the crystal than the electric field.

A5.1.1.4. Poynting vector

The propagation direction of the energy of an electromagnetic wave is given by that of the Poynting vector defined by (see Born & Wolf, 1983)

$$\mathbf{S} = \mathcal{R}(\mathbf{E} \wedge \mathbf{H}^*), \qquad \text{(A5.1.1.8)}$$

where $\mathcal{R}(\,)$ means real part of ( ).

The intensity I of the radiation is equal to the energy crossing unit area per second in the direction normal to that area. It is given by the value of the Poynting vector averaged over a period of time long compared with $1/\nu$:

$$\mathbf{I} = |\mathbf{S}| = c\varepsilon |\mathbf{E}|^2 = c|\mathbf{D}|^2/\varepsilon.$$

References

Adams, B. W. (2004). *Time-dependent Takagi–Taupin eikonal theory of X-ray diffraction in rapidly changing crystal structures*. *Acta Cryst.* A**60**, 120–133.

Authier, A. (1960). *Mise en évidence expérimentale de la double réfraction des rayons X. C. R. Acad. Sci. Paris*, **251**, 2003–2005.

Authier, A. (1961). *Etude de la transmission anomale des rayons X dans des cristaux de silicium*. *Bull. Soc. Fr. Minér. Cristallogr.* **84**, 51–89.

Authier, A. (1970). *Ewald waves in theory and experiment*. *Adv. Struct. Res. Diffr. Methods*, **3**, 1–51.

Authier, A. (1986). *Angular dependence of the absorption induced nodal plane shifts of X-ray stationary waves*. *Acta Cryst.* A**42**, 414–426.

Authier, A. (1989). *X-ray standing waves*. *J. Phys. (Paris)*, **50**, C7–215, C7–224.

Authier, A. (2005). *Dynamical Theory of X-ray Diffraction*. (First printed 2001, revised 2003, 2005.) IUCr Monographs on Crystallography. Oxford University Press.

Authier, A. (2008). *A note on Bragg-case Pendellösung and dispersion surface*. *Acta Cryst.* A**64**, 337–340.

Authier, A. & Balibar, F. (1970). *Création de nouveaux champs d'onde généralisés dus à la présence d'un objet diffractant. II. Cas d'un défaut isolé*. *Acta Cryst.* A**26**, 647–654.

Authier, A., Lagomarsino, S. & Tanner, B. K. (1996). Editors. *X-ray and Neutron Dynamical Diffraction: Theory and Applications*. NATO ASI Series, Series B: Physics, Vol. 357. New York, London: Plenum Press.

Authier, A. & Malgrange, C. (1998). *Diffraction physics*. *Acta Cryst.* A**54**, 806–819.

Batterman, B. W. (1964). *Effect of dynamical diffraction in X-ray fluorescence scattering*. *Phys. Rev. A*, **133**, 759–764.

Batterman, B. W. (1969). *Detection of foreign atom sites by their X-ray fluorescence scattering*. *Phys. Rev. Lett.* **22**, 703–705.

Batterman, B. W. & Bilderback, D. H. (1991). *X-ray monochromators and mirrors*. In *Handbook on Synchrotron Radiation*, Vol. 3, edited by G. Brown & D. E. Moncton, pp. 105–153. Amsterdam: Elsevier Science Publishers BV.

Batterman, B. W. & Cole, H. (1964). *Dynamical diffraction of X-rays by perfect crystals*. *Rev. Mod. Phys.* **36**, 681–717.

Batterman, B. W. & Hildebrandt, G. (1967). *Observation of X-ray Pendellösung fringes in Darwin reflection*. *Phys. Status Solidi*, **23**, K147–K149.

Batterman, B. W. & Hildebrandt, G. (1968). *X-ray Pendellösung fringes in Darwin reflection*. *Acta Cryst.* A**24**, 150–157.

Bedzyk, M. J. (1988). *New trends in X-ray standing waves*. *Nucl. Instrum. Methods A*, **266**, 679–683.

Bonse, U. & Teworte, R. (1980). *Measurement of X-ray scattering factors of Si from the fine structure of Laue case rocking curves*. *J. Appl. Cryst.* **13**, 410–416.

Born, M. & Wolf, E. (1983). *Principles of Optics*, 6th ed. Oxford: Pergamon Press.

Borrmann, G. (1936). *Über die Interferenzen aus Gitterquellen bei Anregung durch Röntgenstrahlen*. *Ann. Phys. (Leipzig)*, **27**, 669–693.

Borrmann, G. (1941). *Über Extinktionsdiagramme der Röntgenstrahlen von Quarz*. *Phys. Z.* **42**, 157–162.

Borrmann, G. (1950). *Die Absorption von Röntgenstrahlen in Fall der Interferenz*. *Z. Phys.* **127**, 297–323.

Borrmann, G. (1954). *Der kleinste Absorption Koeffizient interfierender Röntgenstrahlung*. *Z. Kristallogr.* **106**, 109–121.

Borrmann, G. (1955). *Vierfachbrechung der Röntgenstrahlen durch das ideale Kistallgitter*. *Naturwissenschaften*, **42**, 67–68.

Borrmann, G. (1959). *Röntgenwellenfelder*. *Beit. Phys. Chem. 20 Jahrhunderts*, pp. 262–282. Braunschweig: Vieweg und Sohn.

Bragg, W. L. (1913). *The diffraction of short electromagnetic waves by a crystal*. *Proc. Cambridge Philos. Soc.* **17**, 43–57.

Bragg, W. L., Darwin, C. G. & James, R. W. (1926). *The intensity of reflection of X-rays by crystals*. *Philos. Mag.* **1**, 897–922.

Brümmer, O. & Stephanik, H. (1976). *Dynamische Interferenztheorie*. Leipzig: Akademische Verlagsgesellshaft.

Chang, S.-L. (1987). *Solution to the X-ray phase problem using multiple diffraction – a review*. *Crystallogr. Rev.* **1**, 87–189.

Chang, S.-L. (2004). *X-ray Multiple-Wave Diffraction: Theory and Application*. Springer Series in Solid-State Sciences, Vol. 143. Berlin: Springer-Verlag.

Chang, S.-L., Stetsko, Yu. P. & Lee, Y.-R. (2002). *Quantitative determination of phase for macromolecular crystals using multiple diffraction methods and dynamical theory*. *Z. Kristallogr.* **217**, 662–667.

Chukhovskii, F. N. & Förster, E. (1995). *Time-dependent X-ray Bragg diffraction*. *Acta Cryst.* A**51**, 668–672.

Cowan, P. L., Brennan, S., Jach, T., Bedzyk, M. J. & Materlik, G. (1986). *Observations of the diffraction of evanescent X-rays at a crystal surface*. *Phys. Rev. Lett.* **57**, 2399–2402.

Darwin, C. G. (1914*a*). *The theory of X-ray reflection*. *Philos. Mag.* **27**, 315–333.

Darwin, C. G. (1914*b*). *The theory of X-ray reflection*. *Part II. Philos. Mag.* **27**, 675–690.

Darwin, C. G. (1922). *The reflection of X-rays from imperfect crystals*. *Philos. Mag.* **43**, 800–829.

Ewald, P. P. (1917). *Zur Begründung der Kristalloptik. III. Röntgenstrahlen*. *Ann. Phys. (Leipzig)*, **54**, 519–597.

Ewald, P. P. (1958). *Group velocity and phase velocity in X-ray crystal optics*. *Acta Cryst.* **11**, 888–891.

Fewster, P. F. (1993). *X-ray diffraction from low-dimensional structures*. *Semicond. Sci. Technol.* **8**, 1915–1934.

Fingerland, A. (1971). *Some properties of the single crystal rocking curve in the Bragg case*. *Acta Cryst.* A**27**, 280–284.

Golovchenko, J. A., Patel, J. R., Kaplan, D. R., Cowan, P. L. & Bedzyk, M. J. (1982). *Solution to the surface registration problem using X-ray standing waves*. *Phys. Rev. Lett.* **49**, 560–563.

Graeff, W. (2002*a*). *Short X-ray pulses in a Laue-case crystal*. *J. Synchrotron Rad.* **9**, 82–85.

Graeff, W. (2002*b*). *Time dependence of the polarization of short X-ray pulses after crystal reflection*. *J. Synchrotron Rad.* **9**, 293–297.

Graeff, W. (2004). *Tailoring the time response of a Bragg reflection to short X-ray pulses*. *J. Synchrotron Rad.* **11**, 261–265.

Hart, M. (1981). *Bragg angle measurement and mapping*. *J. Cryst. Growth*, **55**, 409–427.

Hirsch, P. B. & Ramachandran, G. N. (1950). *Intensity of X-ray reflection from perfect and mosaic absorbing crystals*. *Acta Cryst.* **3**, 187–194.

Hümmer, K. & Weckert, E. (1995). *Enantiomorphism and three-beam X-ray diffraction: determination of the absolute structure*. *Acta Cryst.* A**51**, 431–438.

International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

James, R. W. (1950). *The Optical Principles of the Diffraction of X-rays*. London: G. Bell and Sons Ltd.

James, R. W. (1963). *The dynamical theory of X-ray diffraction*. *Solid State Phys.* **15**, 53.

Kato, N. (1952). *Dynamical theory of electron diffraction for a finite polyhedral crystal*. *J. Phys. Soc. Jpn*, **7**, 397–414.

Kato, N. (1955). *Integrated intensities of the diffracted and transmitted X-rays due to ideally perfect crystal*. *J. Phys. Soc. Jpn*, **10**, 46–55.

Kato, N. (1958). *The flow of X-rays and material waves in an ideally perfect single crystal*. *Acta Cryst.* **11**, 885–887.

Kato, N. (1960). *The energy flow of X-rays in an ideally perfect crystal: comparison between theory and experiments*. *Acta Cryst.* **13**, 349–356.

Kato, N. (1961*a*). *A theoretical study of Pendellösung fringes. Part I. General considerations*. *Acta Cryst.* **14**, 526–532.

Kato, N. (1961*b*). *A theoretical study of Pendellösung fringes. Part 2. Detailed discussion based upon a spherical wave theory*. *Acta Cryst.* **14**, 627–636.

Kato, N. (1963). *Pendellösung fringes in distorted crystals. I. Fermat's principle for Bloch waves*. *J. Phys. Soc. Jpn*, **18**, 1785–1791.

Kato, N. (1964*a*). *Pendellösung fringes in distorted crystals. II. Application to two-beam cases*. *J. Phys. Soc. Jpn*, **19**, 67–77.

Kato, N. (1964*b*). *Pendellösung fringes in distorted crystals. III. Application to homogeneously bent crystals*. *J. Phys. Soc. Jpn*, **19**, 971–985.

Kato, N. (1968*a*). *Spherical-wave theory of dynamical X-ray diffraction for absorbing perfect crystals. I. The crystal wave fields*. *J. Appl. Phys.* **39**, 2225–2230.

Kato, N. (1968*b*). *Spherical-wave theory of dynamical X-ray diffraction for absorbing perfect crystals. II. Integrated reflection power*. *J. Appl. Phys.* **39**, 2231–2237.

Kato, N. (1974). *X-ray diffraction*. In *X-ray Diffraction*, edited by L. V. Azaroff, R. Kaplow, N. Kato, R. J. Weiss, A. J. C. Wilson & R. A. Young, pp. 176–438. New York: McGraw-Hill.

Kato, N. & Lang, A. R. (1959). *A study of Pendellösung fringes in X-ray diffraction*. *Acta Cryst.* **12**, 787–794.

Kikuta, S. (1971). *Determination of structure factors of X-rays using half-widths of the Bragg diffraction curves from perfect single crystals*. *Phys. Status Solidi B*, **45**, 333–341.

Kikuta, S. & Kohra, K. (1970). *X-ray collimators using successive asymmetric diffractions and their applications to measurements of diffraction curves. I. General considerations on collimators*. *J. Phys. Soc. Jpn*, **29**, 1322–1328.

Kovalchuk, M. V. & Kohn, V. G. (1986). *X-ray standing waves – a new method of studying the structure of crystals*. *Sov. Phys. Usp.* **29**, 426–446.

Lagomarsino, S. (1996). *X-ray standing wave studies of bulk crystals, thin films and interfaces*. In *X-ray and Neutron Dynamical Diffraction: Theory and Applications*. NATO ASI Series, Series B: Physics, Vol. 357, edited by A. Authier, S. Lagomarsino & B. K. Tanner, pp. 225–234. New York, London: Plenum Press.

Laue, M. von (1931). *Die dynamische Theorie der Röntgenstrahl interferenzen in neuer Form*. *Ergeb. Exakten Naturwiss.* **10**, 133–158.

Laue, M. von (1937). *Helligkeitswechsel längs Kossellinien*. *Ann. Phys.* **28**, 528–532.

Laue, M. von (1952). *Die Energie Strömung bei Röntgenstrahl interferenzen Kristallen*. *Acta Cryst.* **5**, 619–625.

Laue, M. von (1960). *Röntgenstrahl-Interferenzen*. Frankfurt am Main: Akademische Verlagsgesellschaft.

Lefeld-Sosnowska, M. & Malgrange, C. (1968). *Observation of oscillations in rocking curves of the Laue reflected and refracted beams from thin Si single crystals*. *Phys. Status Solidi*, **30**, K23–K25.

Lefeld-Sosnowska, M. & Malgrange, C. (1969). *Experimental evidence of plane-wave rocking curve oscillations*. *Phys. Status Solidi*, **34**, 635–647.

Ludewig, J. (1969). *Debye–Waller factor and anomalous absorption (Ge; 293–5 K)*. *Acta Cryst.* A**25**, 116–118.

Malgrange, C. & Graeff, W. (2003). *Diffraction of short X-ray pulses in the general asymmetric Laue case – an analytic treatment*. *J. Synchrotron Rad.* **10**, 248–254.

Materlik, G. & Zegenhagen, J. (1984). *X-ray standing wave analysis with synchrotron radiation applied for surface and bulk systems*. *Phys. Lett. A*, **104**, 47–50.

Mo, F., Mathiesen, R. H., Alzari, P. M., Lescar, J. & Rasmussen, B. (2002). *Physical estimation of triplet phases from two new proteins*. *Acta Cryst.* D**58**, 1780–1786.

Ohtsuki, Y. H. (1964). *Temperature dependence of X-ray absorption by crystals. I. Photo-electric absorption*. *J. Phys. Soc. Jpn*, **19**, 2285–2292.

Ohtsuki, Y. H. (1965). *Temperature dependence of X-ray absorption by crystals. II. Direct phonon absorption*. *J. Phys. Soc. Jpn*, **20**, 374–380.

Patel, J. R. (1996). *X-ray standing waves*. In *X-ray and Neutron Dynamical Diffraction: Theory and Applications*. NATO ASI Series, Series B: Physics, Vol. 357, edited by A. Authier, S. Lagomarsino & B. K. Tanner, pp. 211–224. New York, London: Plenum Press.

Patel, J. R. & Fontes, E. (1996). *X-ray standing waves: thermal vibration amplitudes at surfaces*. In *X-ray and Neutron Dynamical Diffraction: Theory and Applications*. NATO ASI Series, Series B: Physics, Vol. 357, edited by A. Authier, S. Lagomarsino & B. K. Tanner, pp. 235–248. New York, London: Plenum Press.

Penning, P. & Polder, D. (1961). *Anomalous transmission of X-rays in elastically deformed crystals*. *Philips Res. Rep.* **16**, 419–440.

Pinsker, Z. G. (1978). *Dynamical scattering of X-rays in crystals*. Springer Series in Solid-State Sciences. Berlin: Springer-Verlag.

Prins, J. A. (1930). *Die Reflexion von Röntgenstrahlen an absorbierenden idealen Kristallen*. *Z. Phys.* **63**, 477–493.

Renninger, M. (1955). *Messungen zur Röntgenstrahl-Optik des Idealkristalls. I. Bestätigung der Darwin–Ewald–Prins–Kohler-Kurve*. *Acta Cryst.* **8**, 597–606.

Saka, T., Katagawa, T. & Kato, N. (1973). *The theory of X-ray crystal diffraction for finite polyhedral crystals. III. The Bragg–(Bragg)^m cases*. *Acta Cryst.* A**29**, 192–200.

Shastri, S. D., Zambianchi, P. & Mills, D. M. (2001). *Dynamical diffraction of ultrashort X-ray free-electron laser pulses*. *J. Synchrotron Rad.* **8**, 1131–1135.

Shen, Q. & Wang, J. (2003). *Recursive direct phasing with reference-beam diffraction*. *Acta Cryst.* D**59**, 809–814.

Sondhauss, P. & Wark, J. S. (2003). *Extension of the time-dependent dynamical diffraction theory to 'optical phonon'-type distortions: application to diffraction from coherent acoustic and optical phonons*. *Acta Cryst.* A**59**, 7–13.

Takagi, S. (1962). *Dynamical theory of diffraction applicable to crystals with any kind of small distortion*. *Acta Cryst.* **15**, 1311–1312.

Takagi, S. (1969). *A dynamical theory of diffraction for a distorted crystal*. *J. Phys. Soc. Jpn*, **26**, 1239–1253.

Tanner, B. K. (1976). *X-ray Diffraction Topography*. Oxford: Pergamon Press.

Tanner, B. K. (1990). *High resolution X-ray diffraction and topography for crystal characterization*. *J. Cryst. Growth*, **99**, 1315–1323.

Tanner, B. K. & Bowen, D. K. (1992). *Synchrotron X-radiation topography*. *Mater. Sci. Rep.* **8**, 369–407.

Uragami, T. (1969). *Pendellösung fringes of X-rays in Bragg case*. *J. Phys. Soc. Jpn*, **27**, 147–154.

Uragami, T. (1970). *Pendellösung fringes in a crystal of finite thickness*. *J. Phys. Soc. Jpn*, **28**, 1508–1527.

Vartanyants, I. A. & Koval'chuk, M. V. (2001). *Theory and applications of X-ray standing waves in real crystals*. *Rep. Prog. Phys.* **64**, 1009–1084.

Vartanyants, I. A. & Zegenhagen, J. (1999). *Photoelectric scattering from an X-ray interference field*. *Solid State Commun.* **113**, 299–320.

Wagner, E. H. (1959). *Group velocity and energy (or particle) flow density of waves in a periodic medium*. *Acta Cryst.* **12**, 345–346.

Weckert, E. & Hümmer, K. (1997). *Multiple-beam X-ray diffraction for physical determination of reflection phases and its applications*. *Acta Cryst.* A**53**, 108–143.

Weckert, E., Müller, R., Zellner, J., Zegers, I. & Loris, R. (2002). *Physical measurement of triplet invariants: present state of the experiment, data evaluation and future perspectives*. *Z. Kristallogr.* **217**, 651–661.

Yamazaki, H. & Ishikawa, T. (2002). *Propagation of X-ray coherence for diffraction of perfect crystals*. *J. Appl. Cryst.* **35**, 314–318.

Yamazaki, H. & Ishikawa, T. (2004). *Analysis of the mutual coherence function of X-rays using dynamical diffraction*. *J. Appl. Cryst.* **37**, 48–51.

Zachariasen, W. H. (1945). *Theory of X-ray Diffraction in Crystals*. New York: John Wiley.

Zegenhagen, J. (1993). *Surface structure determination with X-ray standing waves*. *Surf. Sci. Rep.* **18**, 199–271.

Zegenhagen, J., Lyman, P. F., Böhringer, M. & Bedzyk, M. J. (1997). *Discommensurate reconstructions of (111)Si and Ge induced by surface alloying with Cu, Ga and In*. *Phys. Status Solidi B*, **204**, 587–616.

International Tables for Crystallography (2010). Vol. B, Chapter 5.2, pp. 647–653.

5.2. Dynamical theory of electron diffraction

By A. F. Moodie, J. M. Cowley† and P. Goodman†

5.2.1. Introduction

Since electrons are charged, they interact strongly with matter, so that the single scattering approximation has a validity restricted to thin crystals composed of atoms of low atomic number. Further, at energies of above a few tens of keV, the wavelength of the electron is so short that the geometry of two-beam diffraction can be approximated in only small unit cells.

It is therefore necessary to develop a scattering theory specific to electrons and, preferably, applicable to imaging as well as to diffraction. The development, started by Born (1926) and Bethe (1928), and continuing into the present time, is the subject of an extensive literature, which includes reviews [for instance: Howie (1978), Humphreys (1979)] and historical accounts (Goodman, 1981), and is incorporated in Chapter 5.1. Here, an attempt will be made to present only that outline of the main formulations which, it is hoped, will help the nonspecialist in the use of the tables. No attempt will be made to follow the historical development, which has been tortuous and not always logical, but rather to seek the simplest and most transparent approach that is consistent with brevity. Only key points in proofs will be sketched in an attempt to display the nature, rather than the rigorous foundations of the arguments.

5.2.2. The defining equations

No many-body effects have yet been detected in the diffraction of fast electrons, but the velocities lie well within the relativistic region. The one-body Dirac equation would therefore appear to be the appropriate starting point. Fujiwara (1962), using the scattering matrix, carried through the analysis for forward scattering, and found that, to a very good approximation, the effects of spin are negligible, and that the solution is the same as that obtained from the Schrödinger equation provided that the relativistic values for wavelength and mass are used. In effect a Klein–Gordon equation (Messiah, 1965) can be used in electron diffraction (Buxton, 1978) in the form

$$\nabla^2\psi_\mathrm{b} + \frac{8\pi^2 m|e|\varphi}{h^2}\psi_\mathrm{b} + \frac{8\pi^2 m_0|e|W}{h^2}\left(1+\frac{|e|W}{2m_0c^2}\right)\psi_\mathrm{b} = 0.$$

Here, W is the accelerating voltage and φ, the potential in the crystal, is defined as being positive. The relativistic values for mass and wavelength are given by $m = m_0(1 - v^2/c^2)^{-1/2}$, and taking '$e$' now to represent the modulus of the electronic charge, $|e|$,

$$\lambda = h[2m_0eW(1 + eW/2m_0c^2)]^{-1/2},$$

and the wavefunction is labelled with the subscript b in order to indicate that it still includes back scattering, of central importance to LEED (low-energy electron diffraction).

In more compact notation,

$$[\nabla^2 + k^2(1+\varphi/W)]\psi_\mathrm{b} = (\nabla^2 + k^2 + 2k\sigma\varphi)\psi_\mathrm{b} = 0. \quad (5.2.2.1)$$

Here $k = |\mathbf{k}|$ is the scalar wavenumber of magnitude $2\pi/\lambda$, and the interaction constant $\sigma = 2\pi me\lambda/h^2$. This constant is approximately 10^{-3} for 100 kV electrons.

For fast electrons, φ/W is a slowly varying function on a scale of wavelength, and is small compared with unity. The scattering will therefore be peaked about the direction defined by the incident beam, and further simplification is possible, leading to a forward-scattering solution appropriate to HEED (high-energy electron diffraction).

5.2.3. Forward scattering

A great deal of geometric detail can arise at this point and, further, there is no generally accepted method for approximation, the various procedures leading to numerically negligible differences and to expressions of precisely the same form. Detailed descriptions of the geometry are given in the references.

The entrance surface of the specimen, in the form of a plate, is chosen as the x, y plane, and the direction of the incident beam is taken to be close to the z axis. Components of the wavevector are labelled with suffixes in the conventional way; $\mathbf{K}_0 = \mathbf{k}_x + \mathbf{k}_y$ is the transverse wavevector, which will be very small compared to $\mathbf{k}_z$. In this notation, the excitation error for the reflection is given by

$$\zeta_\mathbf{h} = \frac{K_0^2 - |\mathbf{K}_0 + 2\pi\mathbf{h}|^2}{4\pi|\mathbf{k}_z|}.$$

An intuitive method argues that, since $\varphi/W \ll 1$, then the component of the motion along z is little changed by scattering. Hence, making the substitution $\psi_\mathrm{b} = \psi\exp\{ik_z z\}$ and neglecting $\partial^2\psi/\partial z^2$, equation (5.2.2.1) becomes

$$\frac{\partial\psi}{\partial z} = i\left[\frac{1}{2k_z}(\nabla_{x,y}^2 + K_0^2) + \sigma\varphi\right]\psi, \quad (5.2.3.1)$$

where

$$\nabla_{x,y}^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},$$

and $\psi(x, y, 0) = \exp\{i(k_x x + k_y y)\}$.

Equation (5.2.3.1) is of the form of a two-dimensional time-dependent Schrödinger equation, with the z coordinate replacing time. This form has been extensively discussed. For instance, Howie (1966) derived what is essentially this equation using an expansion in Bloch waves, Berry (1971) used a Green function in a detailed and rigorous derivation, and Goodman & Moodie (1974), using methods due to Feynman, derived the equation as the limit of the multislice recurrence relation. A method due to Corones *et al.* (1982) brings out the relationship between the HEED and LEED equations. Equation (5.2.2.1) is cast in the form of a first-order system,

$$\frac{\partial}{\partial z}\begin{pmatrix}\psi_\mathrm{b}\\ \partial\psi_\mathrm{b}/\partial z\end{pmatrix} = \begin{pmatrix}0 & 1\\ -(\nabla_{x,y}^2 + k^2 + 2k\sigma\varphi) & 0\end{pmatrix}\begin{pmatrix}\psi_\mathrm{b}\\ \partial\psi_\mathrm{b}/\partial z\end{pmatrix}.$$

† Deceased.

Copyright © 2010 International Union of Crystallography

A splitting matrix is introduced to separate the wavefunction into the forward and backward components, $\psi_b^{\pm}$, and the fast part of the phase is factored out, so that $\psi_b^{\pm} = \psi^{\pm} \exp\{\pm ik_z z\}$. In the resulting matrix differential equation, the off-diagonal terms are seen to be small for fast electrons, and equation (5.2.2.1) reduces to the pair of equations

$$\frac{\partial \psi^{\pm}}{\partial z} = \pm i\left[\frac{1}{2k_z}(\nabla_{x,y}^2 + K_0^2) + \sigma\varphi\right]\psi^{\pm}. \qquad (5.2.3.2)$$

The equation for $\psi^{\pm}$ is the Lontovich & Fock (1946) parabolic equation.

5.2.4. Evolution operator

Equation (5.2.3.1) is a standard and much studied form, so that many techniques are available for the construction of solutions. One of the most direct utilizes the causal evolution operator. A recent account is given by Gratias & Portier (1983).

In terms of the 'Hamiltonian' of the two-dimensional system,

$$-\mathbf{H}(z) \equiv \frac{1}{2k_z}(\nabla_{x,y}^2 + K_0^2) + \sigma\varphi,$$

the evolution operator $\mathbf{U}(z, z_0)$, defined by $\psi(z) = \mathbf{U}(z, z_0)\psi_0$, satisfies

$$i\frac{\partial}{\partial z}\mathbf{U}(z, z_0) = \mathbf{H}(z)\mathbf{U}(z, z_0), \qquad (5.2.4.1a)$$

or

$$\mathbf{U}(z, z_0) = 1 - i\int_{z_0}^{z} \mathbf{U}(z, z_1)\mathbf{H}(z_1)\,\mathrm{d}z_1. \qquad (5.2.4.1b)$$

5.2.5. Projection approximation – real-space solution

Many of the features of the more general solutions are retained in the practically important projection approximation in which $\varphi(x, y, z)$ is replaced by its projected mean value $\varphi_p(x, y)$, so that the corresponding Hamiltonian $\mathbf{H}_p$ does not depend on z. Equation (5.2.4.1b) can then be solved directly by iteration to give

$$\mathbf{U}_p(z, z_0) = \exp\{-i\mathbf{H}_p(z - z_0)\}, \qquad (5.2.5.1)$$

and the solution may be verified by substitution into equation (5.2.4.1a).

Many of the results of dynamical theory can be obtained by expansion of equation (5.2.5.1) as

$$\mathbf{U}_p \equiv \mathbf{1} - i\mathbf{H}_p(z - z_0) + \frac{i^2}{2!}\mathbf{H}_p^2(z - z_0) - \ldots,$$

followed by the direct evaluation of the differentials. Such expressions can be used, for instance, to explore symmetries in image space.

5.2.6. Semi-reciprocal space

In the derivation of electron-diffraction equations, it is more usual to work in semi-reciprocal space (Tournarie, 1962). This can be achieved by transforming equation (5.2.2.1) with respect to x and y but not with respect to z, to obtain Tournarie's equation

$$\frac{\mathrm{d}^2|U\rangle}{\mathrm{d}z^2} = -\mathbf{M}_b(z)|U\rangle. \qquad (5.2.6.1a)$$

Here $|U\rangle$ is the column vector of scattering amplitudes and $\mathbf{M}_b(z)$ is a matrix, appropriate to LEED, with $\mathbf{k}$ vectors as diagonal elements and Fourier coefficients of the potential as nondiagonal elements.

This equation is factorized in a manner parallel to that used on the real-space equation [equation (5.2.3.1)] (Lynch & Moodie, 1972) to obtain Tournarie's forward-scattering equation

$$\frac{\mathrm{d}|U^{\pm}\rangle}{\mathrm{d}z} = \pm i\mathbf{M}^{\pm}(z)|U^{\pm}\rangle, \qquad (5.2.6.1b)$$

where

$$\mathbf{M}^{\pm}(z) = \pm[\mathbf{K} + (1/2)\mathbf{K}^{-1}V(z)],$$
$$[K_{ij}] = \delta_{ij}K_i,$$

and

$$[V_{ij}] = 2k_z\sum_l V_{i-j}\exp\{-2\pi ilz\},$$

where $V_i \equiv \sigma v_i$ are the scattering coefficients and v_i are the structure amplitudes in volts. In order to simplify the electron-diffraction expression, the third crystallographic index 'l' is taken to represent the periodicity along the z direction.

The double solution involving $\mathbf{M}$ of equation (5.2.6.1b) is of interest in displaying the symmetry of reciprocity, and may be compared with the double solution obtained for the real-space equation [equation (5.2.3.2)]. Normally the $\mathbf{M}^+$ solution will be followed through to give the fast-electron forward-scattering equations appropriate in HEED. $\mathbf{M}^-$, however, represents the equivalent set of equations corresponding to the z reversed reciprocity configuration. Reciprocity solutions will yield diffraction symmetries in the forward direction when coupled with crystal-inverting symmetries (Section 2.5.3).

Once again we set out to solve the forward-scattering equation (5.2.6.1a,b) now in semi-reciprocal space, and define an operator $\mathbf{Q}(z)$ [compare with equation (5.2.4.1a)] such that

$$|U_z\rangle = \mathbf{Q}_z|U_0\rangle \quad \text{with} \quad U_0 = |0\rangle;$$

i.e., $\mathbf{Q}_z$ is an operator that, when acting on the incident wavevector, generates the wavefunction in semi-reciprocal space.

Again, the differential equation can be transformed into an integral equation, and once again this can be iterated. In the projection approximation, with $\mathbf{M}$ independent of z, the solution can be written as

$$\mathbf{Q}_p = \exp\{i\mathbf{M}_p(z - z_0)\}.$$

A typical off-diagonal element is given by $V_{i-j}/\cos\theta_i$, where θ_i is the angle through which the beam is scattered. It is usual in the literature to find that $\cos\theta_i$ has been approximated as unity, since even the most accurate measurements are, so far, in error by much more than this amount.

This expression for $\mathbf{Q}_p$ is Sturkey's (1957) solution, a most useful relation, written explicitly as

$$|U\rangle = \exp\{i\mathbf{M}_\mathrm{p}T\}|0\rangle \tag{5.2.6.2}$$

with T the thickness of the crystal, and $|0\rangle$, the incident state, a column vector with the first entry unity and the rest zero.

$$\mathbf{S} = \exp\{i\mathbf{M}_\mathrm{p}T\}$$

is a unitary matrix, so that in this formulation scattering is described as rotation in Hilbert space.

5.2.7. Two-beam approximation

In the two-beam approximation, as an elementary example, equation (5.2.6.2) takes the form

$$\begin{pmatrix} u_0 \\ u_\mathbf{h} \end{pmatrix} = \exp\left\{ i \begin{pmatrix} 0 & V^*(\mathbf{h}) \\ V(\mathbf{h}) & K_\mathbf{h} \end{pmatrix} T \right\} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.2.7.1}$$

If this expression is expanded directly as a Taylor series, it proves surprisingly difficult to sum. However, the symmetries of Clifford algebra can be exploited by summing in a Pauli basis thus,

$$\begin{aligned} &\exp\left\{ i \begin{pmatrix} 0 & V^*(\mathbf{h}) \\ V(\mathbf{h}) & K_\mathbf{h} \end{pmatrix} T \right\} \\ &\quad = \exp\left\{ i\frac{K_\mathbf{h}T}{2} \right\} \mathbf{E} \exp\left\{ i\left(\frac{K_\mathbf{h}}{2}\boldsymbol{\sigma}_3 + V^\mathrm{R}\boldsymbol{\sigma}_1 - V^\mathrm{I}\boldsymbol{\sigma}_2 \right) T \right\}. \end{aligned}$$

Here, the $\boldsymbol{\sigma}_i$ are the Pauli matrices

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\mathbf{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and V^R, V^I are the real and imaginary parts of the complex scattering coefficients appropriate to a noncentrosymmetric crystal, *i.e.* $V_\mathbf{h} = V^\mathrm{R} + iV^\mathrm{I}$. Expanding,

$$\begin{aligned} &\exp\left\{ i\left(\frac{K_\mathbf{h}}{2}\boldsymbol{\sigma}_3 + V^\mathrm{R}\boldsymbol{\sigma}_1 - V^\mathrm{I}\boldsymbol{\sigma}_2 \right) T \right\} \\ &\quad = \mathbf{E} + i\left(\frac{K_\mathbf{h}}{2}\boldsymbol{\sigma}_3 + V^\mathrm{R}\boldsymbol{\sigma}_1 - V^\mathrm{I}\boldsymbol{\sigma}_2 \right) T \\ &\qquad - \frac{1}{2}\left(\frac{K_\mathbf{h}}{2}\boldsymbol{\sigma}_3 + V^\mathrm{R}\boldsymbol{\sigma}_1 - V^\mathrm{I}\boldsymbol{\sigma}_2 \right)^2 T^2 + \ldots, \end{aligned}$$

using the anti-commuting properties of $\boldsymbol{\sigma}_i$:

$$\left.\begin{aligned} \boldsymbol{\sigma}_i\boldsymbol{\sigma}_j + \boldsymbol{\sigma}_j\boldsymbol{\sigma}_i &= 0 \\ \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i &= 1 \end{aligned}\right\}$$

and putting $[(K_\mathbf{h}/2)^2 + V(\mathbf{h})V^*(\mathbf{h})] = \Omega$, $\mathbf{M}_2 = [(K_\mathbf{h}/2)\boldsymbol{\sigma}_3 + V^\mathrm{R}\boldsymbol{\sigma}_1 - V^\mathrm{I}\boldsymbol{\sigma}_2]$, so that $\mathbf{M}_2^2 = \Omega\mathbf{E}$ and $\mathbf{M}_2^3 = \Omega\mathbf{M}_2$, the powers of the matrix can easily be evaluated. They fall into odd and even series, corresponding to sine and cosine, and the classical two-beam approximation is obtained in the form

$$\mathbf{Q}_2 = \exp\{i(K_\mathbf{h}/2)T\}\mathbf{E}\left[(\cos\Omega^{1/2}T)\mathbf{E} + i\left(\frac{\sin\Omega^{1/2}T}{\Omega^{1/2}} \right)\mathbf{M}_2 \right]. \tag{5.2.7.2}$$

This result was first obtained by Blackman (1939), using Bethe's dispersion formulation. Ewald and, independently, Darwin, each with different techniques, had, in establishing the theoretical foundations for X-ray diffraction, obtained analogous results (see Section 5.1.3).

The two-beam approximation, despite its simplicity, exemplifies some of the characteristics of the full dynamical theory, for instance in the coupling between beams. As Ewald pointed out, a formal analogy can be found in classical mechanics with the motion of coupled pendulums. In addition, the functional form $(\sin ax)/x$, deriving from the shape function of the crystal emerges, as it does, albeit less obviously, in the N-beam theory.

This derivation of equation (5.2.7.2) exhibits two-beam diffraction as a typical two-level system having analogies with, for instance, lasers and nuclear magnetic resonance and exhibiting the symmetries of the special unitary group SU(2) (Gilmore, 1974).

5.2.8. Eigenvalue approach

In terms of the eigenvalues and eigenvectors, defined by

$$\mathbf{H}_\mathrm{p}|j\rangle = \gamma_j|j\rangle,$$

the evolution operator can be written as

$$\mathbf{U}(z, z_0) = \int |j\rangle \exp\{\gamma_j(z - z_0)\}\langle j|\,\mathrm{d}j.$$

This integration becomes a summation over discrete eigen states when an infinitely periodic potential is considered.

Despite the early developments by Bethe (1928), an N-beam expression for a transmitted wavefunction in terms of the eigenvalues and eigenvectors of the problem was not obtained until Fujimoto (1959) derived the expression

$$U_\mathbf{h} = \sum_j \psi_0^{j*}\psi_\mathbf{h}^j \exp\{-i2\pi\gamma_j T\}, \tag{5.2.8.1}$$

where $\psi_\mathbf{h}^j$ is the h component of the j eigenvector with eigenvalue γ_j.

This expression can now be related to those obtained in the other formulations. For example, Sylvester's theorem (Frazer *et al.*, 1963) in the form

$$f(\mathbf{M}) = \sum_j \mathbf{A}_j f(\gamma_j)$$

when applied to Sturkey's solution yields

$$\boldsymbol{\Phi}_\mathbf{h} = \exp(i\mathbf{M}_\mathrm{p}z) = \sum \mathbf{P}_j \exp\{i2\pi\gamma_j z\}$$

(Kainuma, 1968; Hurley *et al.*, 1978). Here, the $\mathbf{P}_j$ are projection operators, typically of the form

$$\mathbf{P}_j = \prod_{n\neq j} \frac{(\mathbf{M}_\mathrm{p} - \mathbf{E}\gamma_n)}{\gamma_j - \gamma_n}.$$

On changing to a lattice basis, these transform to $\psi_0^{j*}\psi_{\mathbf{h}}^{j}$.

Alternatively, the semi-reciprocal differential equation can be uncoupled by diagonalizing $\mathbf{M}_{\rm p}$ (Goodman & Moodie, 1974), a process which involves the solution of the characteristic equation

$$|\mathbf{M}_{\rm p} - \gamma_j \mathbf{E}| = 0. \tag{5.2.8.2}$$

5.2.9. Translational invariance

An important result deriving from Bethe's initial analysis, and not made explicit in the preceding formulations, is that the fundamental symmetry of a crystal, namely translational invariance, by itself imposes a specific form on wavefunctions satisfying Schrödinger's equation.

Suppose that, in a one-dimensional description, the potential in a Hamiltonian $\mathbf{H}_t(x)$ is periodic, with period t. Then,

$$\varphi(x + t) = \varphi(x)$$

and

$$\mathbf{H}_t\psi(x) = \mathbf{E}\psi(x).$$

Now define a translation operator

$$\mathbf{\Gamma} f(x) = f(x + t),$$

for arbitrary $f(x)$. Then, since $\mathbf{\Gamma}\varphi(x) = \varphi(x)$, and ∇^2 is invariant under translation,

$$\mathbf{\Gamma}\mathbf{H}_t(x) = \mathbf{H}_t(x)$$

and

$$\mathbf{\Gamma}\mathbf{H}_t(x)\psi(x) = \mathbf{H}_t(x + t)\psi(x + t) = \mathbf{H}_t(x)\mathbf{\Gamma}\psi(x).$$

Thus, the translation operator and the Hamiltonian commute, and therefore have the same eigenfunctions (but not of course the same eigenvalues), *i.e.*

$$\mathbf{\Gamma}\psi(x) = \alpha\psi(x).$$

This is a simpler equation to deal with than that involving the Hamiltonian, since raising the operator to an arbitrary power simply increments the argument

$$\mathbf{\Gamma}^m\psi(x) = \psi(x + mt) = \alpha^m\psi(x).$$

But $\psi(x)$ is bounded over the entire range of its argument, positive and negative, so that $|\alpha| = 1$, and α must be of the form $\exp\{i2\pi kt\}$.

Thus, $\psi(x + t) = \mathbf{\Gamma}\psi(x) = \exp\{i2\pi kt\}\psi(x)$, for which the solution is

$$\psi(x) = \exp\{i2\pi kt\}q(x)$$

with $q(x + t) = q(x)$.

This is the result derived independently by Bethe and Bloch. Functions of this form constitute bases for the translation group, and are generally known as Bloch functions. When extended in a direct fashion into three dimensions, functions of this form ultimately embody the symmetries of the Bravais lattice; *i.e.* Bloch functions are the irreducible representations of the translational component of the space group.

5.2.10. Bloch-wave formulations

In developing the theory from the beginning by eigenvalue techniques, it is usual to invoke the periodicity of the crystal in order to show that the solutions to the wave equation for a given wavevector $\mathbf{k}$ are Bloch waves of the form

$$\psi = C(\mathbf{r})\exp\{i\mathbf{k}\cdot\mathbf{r}\},$$

where $C(\mathbf{r})$ has the periodicity of the lattice, and hence may be expanded in a Fourier series to give

$$\psi = \sum_{\mathbf{h}} C_{\mathbf{h}}(\mathbf{k})\exp\{i(\mathbf{k} + 2\pi\mathbf{h})\cdot\mathbf{r}\}. \tag{5.2.10.1}$$

The $C_{\mathbf{h}}(\mathbf{k})$ are determined by equations of consistency obtained by substitution of equation (5.2.10.1) into the wave equation.

If N terms are selected in equation (5.2.10.1) there will be N Bloch waves where wavevectors differ only in their components normal to the crystal surface, and the total wavefunction will consist of a linear combination of these Bloch waves. The problem is now reduced to the problem of equation (5.2.8.2).

The development of solutions for particular geometries follows that for the X-ray case, Chapter 5.1, with the notable differences that:

(1) The two-beam solution is not adequate except as a first approximation for particular orientations of crystals having small unit cells and for accelerating voltages not greater than about 100 keV. In general, many-beam solutions must be sought.

(2) For transmission HEED, the scattering angles are sufficiently small to allow the use of a small-angle forward-scattering approximation.

(3) Polarization effects are negligible except for very low energy electrons.

Humphreys (1979) compares the action of the crystal, in the Bloch-wave formalism, with that of an interferometer, the incident beam being partitioned into a set of Bloch waves of different wavevectors. 'As each Bloch wave propagates it becomes out of phase with its neighbours (due to its different wavevector). Hence interference occurs. For example, if the crystal thickness varies, interference fringes known as thickness fringes are formed.' For the two-beam case, these are the fringes of the pendulum solution referred to previously.

5.2.11. Dispersion surfaces

One of the important constructs of the Bloch-wave formalism is the dispersion surface, a plot of the permitted values of the z component of a Bloch wavevector against the component of the incident wavevector parallel to the crystal surface. The curve for a particular Bloch wave is called a branch. Thus, for fast electrons, the two-beam approximation has two branches, one for each eigenvalue, and the N-beam approximation has N.

A detailed treatment of the extensive and powerful theory that has grown from Bethe's initial paper is to be found, for example, in Hirsch *et al.* (1965). Apart from its fundamental importance as a theoretical tool, this formulation provides the basis for one of the most commonly used numerical techniques, the essential step being the estimation of the eigenvalues from equation (5.2.8.2) [see *IT* C (2004, Section 4.3.6.2)].

5.2.12. Multislice

Multislice derives from a formulation that generates a solution in the form of a Born series (Cowley & Moodie, 1962). The crystal is treated as a series of scattering planes on to which the potential from the slice between z and $z + \Delta z$ is projected, separated by vacuum gaps Δz, not necessarily corresponding to any planes or spacings of the material structure. The phase change in the electron beam produced by passage through a slice is given by

$$q = \exp\left\{-i\sigma \int_{z_1}^{z_1+\Delta z} \varphi(x, y, z)\,\mathrm{d}z\right\},$$

and the phase distribution in the x, y plane resulting from propagation between slices is given by

$$p = \exp\left\{\frac{ik(x^2 + y^2)}{2\Delta z}\right\},$$

where the wavefront has been approximated by a paraboloid. Thus, the wavefunction for the $(n + 1)$th slice is given by

$$\begin{aligned}\psi_{n+1} &= \left[\psi_n * \exp\left\{\frac{ik(x^2 + y^2)}{2\Delta z}\right\}\right] \exp\{-i\sigma\varphi_{n+1}\} \\ &= [\psi_n * p]q, \end{aligned} \tag{5.2.12.1}$$

where $*$ is the convolution operator (Cowley, 1981).

This equation can be regarded as the finite difference form of the Schrödinger equation derived by Feynman's (1948) method. The calculation need be correct only to first order in Δz. Writing the convolution in equation (5.2.12.1) explicitly, and expanding in a Taylor series, the integrals can be evaluated to yield equation (5.2.3.1) (Goodman & Moodie, 1974).

If equation (5.2.12.1) is Fourier transformed with respect to x and y, the resulting recurrence relation is of the form

$$U_{n+1} = [U_n P] * Q_n, \tag{5.2.12.2}$$

where P and Q are obtained by Fourier transforming p and q above. This form is convenient for numerical work since, for a perfect crystal, it is: discrete, as distinct from equation (5.2.12.1) which is continuous in the variables [see *IT* C (2004, Section 4.3.6.1)]; numerically stable at least up to 5000 beams; fast; and only requires a computer memory proportional to the number of beams (Goodman & Moodie, 1974).

5.2.13. Born series

In the impulse limit of equation (5.2.12.2), the integrals can be evaluated to give the Born series (Cowley & Moodie, 1957)

$$U(h, k) = \sum_n U_n(h, k),$$

where

$$\begin{aligned} U_n(h, k) = \sum_l \sum_{h_1 k_1 l_1} \cdots \sum_{h_{n-1} k_{n-1} l_{n-1}} & i^n V(h_1, k_1, l_1) \\ & \ldots V\left(h - \sum_{r=1}^{n-1} h_r, k - \sum_{r=1}^{n-1} k_r, l - \sum_{r=1}^{n-i} l_r\right) \\ & \times [\exp\{-2\pi i\zeta T\}/(2\pi i)^n] \\ & \times \Big(\exp\{i\pi\zeta T\}(\sin \pi\zeta T/\zeta)\big[(\zeta - \zeta_1)\ldots(\zeta - \zeta_{n-1})\big]^{-1} \\ & + \sum_{m=1}^{n-1} \exp\{i\pi\zeta_m T\}(\sin \pi\zeta_m T/\zeta_m)\big[(\zeta_m - \zeta_1) \\ & \ldots(\zeta_m - \zeta_{m-1})(\zeta_m - \zeta_{m+1})\ldots(\zeta_m - \zeta)\big]^{-1}\Big) \end{aligned} \tag{5.2.13.1a}$$

and where n is the order of interaction. Here ζ is the excitation error of the reflection with index h, k, and ζ_i are the excitation errors for the reflections with indices h_i, k_i, l_i. Thus each constituent process may be represented by a diagram, starting on the origin of reciprocal space, possibly looped, and ending on the point with coordinates (h, k).

This solution can also be obtained by iteration of the Green-function integral equation, the integrals being evaluated by means of suitably chosen contours on the complex k_z plane (Fujiwara, 1959), as well as by expansion of the scattering matrix (Fujimoto, 1959).

Clearly, two or more of the ζ_i will, in general, be equal in nearly all of the terms in equation (5.2.13.1*a*). Confluence is, however, readily described, the divided differences of arbitrary order transforming into differentials of the same order (Moodie, 1972).

The physical picture that emerges from equation (5.2.13.1*a*) is that of n-fold scattering, the initial wave being turned through $n - 1$ intermediate states, processes that can be presented by scattering diagrams in reciprocal space (Gjønnes & Moodie, 1965).

For a given scattering vector, constituent functions are evaluated for all possible paths in three dimensions, and those functions are then summed over l. There are therefore two distinct processes by which upper-layer lines can perturb wavefunctions in the zone, namely: by scattering out of the zone and then back in; and by intrusion of the effective shape function from another zone, the latter process being already operative in the first Born, or kinematical approximation.

The constituent functions to be evaluated can be transformed into many forms. One of the more readily described is that which assigns to each diagram an effective dynamical shape function. If there are no loops in the diagram of order n, this effective shape function is the $(n + 1)$th divided difference of the constituent phase-shifted kinematical shape transforms. For general diagrams, divided differences in loops are replaced by the corresponding differentials. The resulting function is multiplied by the convolution of the contributing structure amplitudes and diagrams of all orders summed (Moodie, 1972).

While scattering diagrams have no utility in numerical work, they find application in the analysis of symmetries, for instance in the determination of the presence or absence of a centre of inversion [for a recent treatment, see Moodie & Whitfield (1995)] and in the detection of screw axes and glide planes (Gjønnes & Moodie, 1965). Methods for the direct determination of all space groups are described by Goodman (1975) and by Tanaka *et al.* (1983) (see Section 2.5.3).

Equation (5.2.13.1*a*) can be rewritten in a form particularly suited to the classification of approximations, and to describing the underlying symmetry of the formulation. The equation is written for compactness as

$$U_n(h) = E_n(h)Z_n(\zeta),$$

so that $E_n(h)$ depends only on crystal structure and $Z_n(\zeta)$ only on diffraction geometry. A transformation (Cowley & Moodie, 1962) involving bialternants leads to

$$U_n = \sum_{r=0}^{\infty} E_n(h)\big[(2\pi iT)^{n+r}/(n+r)!\big]h_r(\zeta, \zeta_1 \ldots \zeta_{n-1}), \quad (5.2.13.1b)$$

where h_r is the complete homogeneous symmetric polynomial function of n variables of order r.

Upper-layer-line effects can, of course, be calculated in any of the formulations.

5.2.14. Approximations

So far, only the familiar first Born and two-beam approximations and the projection approximation have been mentioned. Several others, however, have a considerable utility.

A high-voltage limit can be calculated in standard fashion to give

$$U_{\rm HVL}(h, k) = \mathscr{F} \exp\left\{-i\sigma_c \int_0^T \varphi(x, y, z)\, {\rm d}z\right\}, \quad (5.2.14.1)$$

where $\mathscr{F}$ is the Fourier transform operator, and $\sigma_c = 2\pi m_0 e\lambda_c/h^2$ with $\lambda_c = (h/m_0 c)$, the Compton wavelength. The phase-grating approximation, which finds application in electron microscopy, involves the assumption that equation (5.2.14.1) has some range of validity when σ_c is replaced by σ. This is equivalent to ignoring the curvature of the Ewald sphere and can therefore apply to thin crystals [see Section 2.5.2 and *IT* C (2004, Section 4.3.8)].

Approximations that involve curtailing the number of beams evidently have a range of validity that depends on the size of the unit cell. The most explored case is that of three-beam interactions. Kambe (1957) has demonstrated that phase information can be obtained from the diffraction data; Gjønnes & Høier (1971) analysed the confluent case, and Hurley & Moodie (1980) have given an explicit inversion for the centrosymmetric case. Analyses of the symmetry of the defining differential equation, and of the geometry of the noncentrosymmetric case, have been given by Moodie *et al.* (1996, 1998).

Niehrs and his co-workers (*e.g.* Blume, 1966) have shown that, at or near zones, effective two-beam conditions can sometimes obtain, in that, for instance, the central beam and six equidistant beams of equal structure amplitude can exhibit two-beam behaviour when the excitation errors are equal. Group-theoretical treatments have been given by Fukuhara (1966) and by Kogiso & Takahashi (1977). Explicit reductions for all admissible noncentrosymmetric space groups have been obtained by Moodie & Whitfield (1994). Extensions of such results have application in the interpretation of lattice images and convergent-beam patterns.

The approximations near the classical limit have been extensively explored [for instance, see Berry (1971)] but channelling has effectively become a separate subject and cannot be discussed here.

References

Berry, M. V. (1971). *Diffraction in crystals at high voltages. J. Phys. C*, **4**, 697–722.

Bethe, H. A. (1928). *Theorie der Beugung von Elektronen an Kristallen. Ann. Phys. (Leipzig)*, **87**, 55–129.

Blackman, M. (1939). *Intensities of electron diffraction rings. Proc. Phys. Soc. London Sect. A*, **173**, 68–72.

Blume, J. (1966). *Die Kantenstreifung im Elektronen-Mikroskopischen Bild Wurfelformiger MgO Kristalle bei Durchstrahlung im Richtung der Raumdiagonal. Z. Phys.* **191**, 248–272.

Born, M. (1926). *Quantenmechanik der Stossvorgange. Z. Phys.* **38**, 803–826.

Buxton, B. (1978). *Graduate Lecture-Course Notes: Dynamical Diffraction Theory.* Cambridge University, England.

Corones, J., De Facio, B. & Kreuger, R. J. (1982). *Parabolic approximations to the time-independent elastic wave equation. J. Math. Phys.* **23**, 577–586.

Cowley, J. M. (1981). *Diffraction Physics*, pp. 26–30. Amsterdam: North-Holland.

Cowley, J. M. & Moodie, A. F. (1957). *The scattering of electrons by atoms and crystals. I. A new theoretical approach. Acta Cryst.* **10**, 609–619.

Cowley, J. M. & Moodie, A. F. (1962). *The scattering of electrons by thin crystals. J. Phys. Soc. Jpn*, **17**, Suppl. B11, 86–91.

Feynman, R. (1948). *Space–time approach to non-relativistic quantum mechanics. Rev. Mod. Phys.* **201**, 367–387.

Frazer, R. A., Duncan, W. J. & Collar, A. R. (1963). *Elementary Matrices*, pp. 78–79. Cambridge University Press.

Fujimoto, F. (1959). *Dynamical theory of electron diffraction in the Laue case. J. Phys. Soc. Jpn*, **14**, 1558–1568.

Fujiwara, K. (1959). *Application of higher order Born approximation to multiple elastic scattering of electrons by crystals. J. Phys. Soc. Jpn*, **14**, 1513–1524.

Fujiwara, K. (1962). *Relativistic dynamical theory of electron diffraction. J. Phys. Soc. Jpn*, **17**, Suppl. B11, 118–123.

Fukuhara, A. (1966). *Many-ray approximations in the dynamical theory of electron diffraction. J. Phys. Soc. Jpn*, **21**, 2645–2662.

Gilmore, R. (1974). *Lie Groups, Lie Algebras, and Some of Their Applications.* New York: Wiley–Interscience.

Gjønnes, J. & Høier, R. (1971). *The application of non-systematic many-beam dynamic effects to structure-factor determination. Acta Cryst.* A**27**, 313–316.

Gjønnes, J. & Moodie, A. F. (1965). *Extinction conditions in the dynamic theory of electron diffraction. Acta Cryst.* **19**, 65–67.

Goodman, P. (1975). *A practical method of three-dimensional space-group analysis using convergent-beam electron diffraction. Acta Cryst.* A**31**, 804–810.

Goodman, P. (1981). Editor. *Fifty Years of Electron Diffraction.* Dordrecht: Kluwer Academic Publishers.

Goodman, P. & Moodie, A. F. (1974). *Numerical evaluation of N-beam wave functions in electron scattering by the multislice method. Acta Cryst.* A**30**, 280–290.

Gratias, D. & Portier, R. (1983). *Time-like perturbation method in high energy electron diffraction. Acta Cryst.* A**39**, 576–584.

Hirsch, P. B., Howie, A., Nicholson, R. B., Pashley, D. W. & Whelan, M. J. (1965). *Electron Microscopy of Thin Crystals.* London: Butterworths.

Howie, A. (1966). *Diffraction channelling of fast electrons and positrons in crystals. Philos. Mag.* **14**, 223–237.

Howie, A. (1978). In *Electron Diffraction 1927–1977*, edited by P. J. Dobson, J. B. Pendry & C. J. Humphreys, pp. 1–12. *Inst. Phys. Conf. Ser.* No. 41. Bristol/London: Institute of Physics.

Humphreys, C. J. (1979). *The scattering of fast electrons by crystals. Rep. Prog. Phys.* **42**, 1825–1887.

Hurley, A. C., Johnson, A. W. S., Moodie, A. F., Rez, P. & Sellar, J. R. (1978). *Algebraic approaches to N-beam theory.* In *Electron Diffraction 1927–1977*, edited by P. J. Dobson, J. B. Pendry & C. J. Humphreys, pp. 34–40. *Inst. Phys. Conf. Ser.* No. 41. Bristol/London: Institute of Physics.

Hurley, A. C. & Moodie, A. F. (1980). *The inversion of the three-beam intensities for scalar scattering by a general centrosymmetric crystal. Acta Cryst.* A**36**, 737–738.

International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

Kainuma, Y. (1968). *Averaged intensities in the many beam dynamical theory of electron diffraction. Part I. J. Phys. Soc. Jpn*, **25**, 498–510.

Kambe, K. (1957). *Study of simultaneous reflection in electron diffraction by crystal. J. Phys. Soc. Jpn*, **12**, 13–31.

Kogiso, M. & Takahashi, H. (1977). *Group-theoretical method in the many-beam theory of electron diffraction. J. Phys. Soc. Jpn*, **42**, 223–229.

Lontovitch, M. & Fock, R. (1946). *Solution of the problem of propagation of electromagnetic waves along the Earth's surface by the method of parabolic equation.* (Translated from Russian by J. Smorodinsky.) *J. Phys.* **10**, 13–24.

Lynch, D. F. & Moodie, A. F. (1972). *Numerical evaluation of low energy electron diffraction intensities. Surf. Sci.* **32**, 422–438.

Messiah, A. (1965). *Quantum Mechanics*, Vol. II, pp. 884–888. Amsterdam: North-Holland.

Moodie, A. F. (1972). *Reciprocity and shape functions in multiple scattering diagrams. Z. Naturforsch. Teil A*, **27**, 437–440.

Moodie, A. F., Etheridge, J. & Humphreys, C. J. (1996). *The symmetry of three-beam scattering equations: inversion of three-beam diffraction patterns from centrosymmetric crystals. Acta Cryst.* A**52**, 596–605.

Moodie, A. F., Etheridge, J. & Humphreys, C. J. (1998). *The Coulomb interaction and the direct measurement of structural phase*. In *The Electron*, edited by A. Kirkland & P. Brown, pp. 235–246. IOM Communications Ltd. London: The Institute of Materials.

Moodie, A. F. & Whitfield, H. J. (1994). *The reduction of N-beam scattering from noncentrosymmetric crystals to two-beam form*. *Acta Cryst.* A**50**, 730–736.

Moodie, A. F. & Whitfield, H. J. (1995). *Friedel's law and non-centrosymmetric space groups*. *Acta Cryst.* A**51**, 198–201.

Sturkey, L. (1957). *The use of electron-diffraction intensities in structure determination. Acta Cryst.* **10**, 858–859.

Tanaka, M., Sekii, H. & Nagasawa, T. (1983). *Space-group determination by dynamic extinction in convergent-beam electron diffraction*. *Acta Cryst.* A**39**, 825–837.

Tournarie, M. (1962). *Recent developments of the matrical and semi-reciprocal formulation on the field of dynamical theory. J. Phys. Soc. Jpn*, **17**, Suppl. B11, 98–100.

International Tables for Crystallography (2010). Vol. B, Chapter 5.3, pp. 654–664.

5.3. Dynamical theory of neutron diffraction

By M. Schlenker and J.-P. Guigay

5.3.1. Introduction

Neutron and X-ray scattering are quite similar both in the geometry of scattering and in the orders of magnitude of the basic quantities. When the neutron spin is neglected, *i.e.* when dealing with scattering by perfect nonmagnetic crystals, the formalism and the results of the dynamical theory of X-ray scattering can be very simply transferred to the case of neutrons (Section 5.3.2). Additional features of the neutron case are related to the neutron spin and appear in diffraction by magnetic crystals (Section 5.3.3). The low intensities available, coupled with the low absorption of neutrons by most materials, make it both necessary and possible to use large samples in standard diffraction work. The effect of extinction in crystals that are neither small nor bad enough to be amenable to the kinematical approximation is therefore very important in the neutron case, and will be discussed in Section 5.3.4 together with the effect of crystal distortion. Additional possibilities arise in the neutron case because the neutrons can be manipulated from outside through applied fields (Section 5.3.5). Reasonably extensive tests of the predictions of the dynamical theory of neutron diffraction have been performed, with the handicap of the very low intensities of neutron beams as compared with X-rays: these are described in Section 5.3.6. Finally, the applications of the dynamical theory in the neutron case, and in particular neutron interferometry, are reviewed in Section 5.3.7.

5.3.2. Comparison between X-rays and neutrons with spin neglected

5.3.2.1. The neutron and its interactions

An excellent introductory presentation of the production, properties and scattering properties of neutrons is available (Scherm & Fåk, 1993, and other papers in the same book). A stimulating review on neutron optics, including diffraction by perfect crystals, has been written by Klein & Werner (1983). X-rays and neutrons are compared in terms of the basic quantities in Table 4.1.3.1 of *IT* C (2004), where Chapter 4.4 is devoted to neutron techniques.

The neutron is a massive particle for which the values relevant to diffraction are: no electric charge, rest mass $m = 1.675 \times 10^{-27}$ kg, angular momentum eigenvalues along a given direction $\pm\hbar/2$ (spin $\frac{1}{2}$) and a magnetic moment of -1.913 nuclear magneton, meaning that its component along a quantization direction z can take eigenvalues $\mu_z = \mp 0.996 \times 10^{-26}$ A m^2. The de Broglie wavelength is $\lambda = h/p$ where h is Planck's constant ($h = 2\pi\hbar = 6.625 \times 10^{-34}$ J s) and p is the linear momentum; $p = mv$ in the nonrelativistic approximation, which always applies in the context of this chapter, v being the neutron's velocity. The neutron's wavelength, λ, and kinetic energy, E_c, are thus related by $\lambda = h/(2mE_c)^{1/2}$, or, in practical units, $\lambda\,[\text{Å}] = 9.05/(E_c\,[\text{meV}])^{1/2}$. Thus, to be of interest for diffraction by materials, neutrons should have kinetic energies in the range 10^0 to 10^2 meV. In terms of the velocity, $\lambda\ [\text{Å}] = 3.956/(v\ [\text{km s}^{-1}])$.

Neutron beams are produced by nuclear reactors or by spallation sources, usually pulsed. In either case they initially have an energy in the MeV range, and have to lose most of it before they can be used. The moderation process involves inelastic interactions with materials. It results in statistical distributions of energy, hence of velocity, close to the Maxwell distribution characteristic of the temperature T of the moderator. Frequently used moderators are liquid deuterium (D_2, *i.e.* $^2H_2$) at 25 K, heavy water (D_2O) at room temperature and graphite allowed to heat up to 2400 K; the corresponding neutron distributions are termed cold, thermal and hot, respectively.

The interaction of a neutron with an atom is usually described in terms of scattering lengths or of scattering cross sections. The main contribution corresponding to the nuclear interaction is related to the strong force. The interaction with the magnetic field created by atoms with electronic magnetic moments is comparable in magnitude to the nuclear term.

5.3.2.2. Scattering lengths and refractive index

The elastic scattering amplitude for scattering vector $\mathbf{s}$, $f(\mathbf{s})$, is defined by the wave scattered by an object placed at the origin when the incident plane wave is $\Psi_i = A\exp[i(\mathbf{k}_0\cdot\mathbf{r} - \omega t)]$, written as $\Psi_s = A[f(\mathbf{s})/r]\exp[i(kr - \omega t)]$ with $k = |\mathbf{k}_0| = |\mathbf{k}_0 + \mathbf{s}| = 2\pi/\lambda$. In the case of the strong-force interaction with nuclei, the latter can be considered as point scatterers because the interaction range is very small, hence the scattering amplitude is isotropic (independent of the direction of $\mathbf{s}$). It is also independent of λ except in the vicinity of resonances. It is conventionally written as $-b$ so that most values of b, called the scattering length, are positive. A table of experimentally measured values of the scattering lengths b is given in *IT* C for the elements in their natural form as well as for many individual isotopes. It is apparent that the typical order of magnitude is the fm (femtometre, *i.e.* 10^{-15} m, or fermi), that there is no systematic variation with atomic number and that different isotopes have very different scattering lengths, including different signs. The first remark implies that scattering amplitudes of X-rays and of neutrons have comparable magnitudes, because the characteristic length for X-ray scattering (the scattering amplitude for forward scattering by one free electron) is $R = 2.8$ fm, the classical electron radius. The second and third points explain the importance of neutrons in structural crystallography, in diffuse scattering and in small-angle scattering. Scattering of neutrons by condensed matter implies the use of the bound scattering lengths, as tabulated in *IT* C. The 'free' scattering length, used in some presentations, is obtained by multiplying the bound scattering lengths by $A/(A+1)$, where A is the mass of the nucleus in atomic units.

A description in terms of an interaction potential is possible using the Fermi pseudo-potential, which in the case of the nuclear interaction with a nucleus at $\mathbf{r}_0$ can be written as $V(\mathbf{r}) = (h^2/2\pi m)b\delta(\mathbf{r} - \mathbf{r}_0)$, where δ denotes the three-dimensional Dirac distribution.

Refraction of neutrons at an interface can be conveniently described by assigning a refractive index to the material, such that the wavenumber in the material, k, is related to that in a vacuum, k_0, by $k = nk_0$. Here

$$n = \left(1 - \frac{\lambda^2}{\pi V}\sum_i b_i\right)^{1/2},$$

where the sum is over the nuclei contained in volume V. With typical values, n is very close to 1 and $1 - n = (\lambda^2/2\pi V)\sum_i b_i$ is

Copyright © 2010 International Union of Crystallography

typically of the order of 10^{-5}. This small value, in the same range as for X-rays, gives a feeling for the order of magnitude of key quantities of the dynamical theory, in particular the Darwin width 2δ as discussed in Chapter 5.1. It also makes total external reflection possible on materials for which $\sum_i b_i > 0$: this is the basis for the neutron guide tubes now installed in most research reactors, as well as for reflectometry.

The notations prevailing in X-ray and in neutron crystallography are slightly different, and the correspondence is very simple: X-ray atomic scattering factors and structure factors are numbers. When multiplied by R, the classical electron radius, they become entirely equivalent to the corresponding quantities in neutron usage, which are lengths. It should be noted that the presence of different isotopes and the effect of nuclear spin (disordered except under very special conditions) give rise to incoherent elastic neutron scattering, which has no equivalent in the X-ray case. The scattering length corresponding to R times the atomic scattering factor for X-rays is therefore the coherent scattering length, $b_{\rm coh}$, obtained by averaging the scattering length over the nuclear spin state and isotope distribution.

5.3.2.3. Absorption

Neutron absorption is related to a nuclear reaction in which the neutron combines with the absorbing nucleus to form a compound nucleus, usually in a metastable state which then decays. The scattering length describing this resonance scattering process depends on the neutron energy and contains an imaginary part associated with absorption in complete analogy with the imaginary part of the dispersion correction for the X-ray atomic scattering factors. The energies of the resonances are usually far above those of interest for crystallography, and the linear absorption coefficient varies approximately as $1/v$ or λ. It is important to note that, except for a very few cases (notably ^{3}He, ^{6}Li, ^{10}B, In, Cd, Gd), the absorption of neutrons is very small compared with that of X-rays, and even more so compared with that of electrons, and can be neglected to a first approximation.

5.3.2.4. Differences between neutron and X-ray scattering

There are major differences in the experimental aspects of neutron and X-ray scattering. Neutrons are only available in large facilities, where allocation of beam time to users is made on the basis of applications, and where admittance is restricted because of the hazards which nuclear technology can present in the hands of ill-intentioned users. Because of the radiation shielding necessary, as well as the large size of neutron detectors, neutron-scattering instrumentation is much bulkier than that for X-rays. Neutron beams are in some aspects similar to synchrotron radiation, in particular because in both cases the beams are initially 'white' and for most applications have to be monochromated. There is, however, a huge difference in the order of magnitudes of the intensities. Neutron beams are weak in comparison with laboratory X-ray sources, and weaker by many orders of magnitude than synchrotron radiation. Also, the beam sources are large in the case of neutrons, since they are essentially the moderators, whereas the source is very small in the case of synchrotron radiation, and this difference again increases the ratio of the brilliances in favour of X-rays. This encourages the use of large specimens in all neutron-scattering work, and makes the extinction problem more important than for X-rays. Furthermore, many experiments that are quick using X-rays become very slow, and give rise to impaired resolution, in the neutron case.

There are also at least two additional aspects of neutron scattering in comparison with X-ray scattering, apart from the effect of the magnetic moment associated with the intrinsic (spin) angular momentum of the neutron. On the one hand, the small velocity of neutrons, compared with the velocity of light, makes time-of-flight measurements possible, both in standard neutron diffraction and in investigations of perfect crystals. Because this velocity is of the same order of magnitude as that of ultrasound, the effect of ultrasonic excitation on neutron diffraction is slightly different from that in the X-ray case. On the other hand, the fact that neutrons have mass and a magnetic moment implies that they can be affected by external fields, such as gravity and magnetic fields, both during their propagation in air or in a vacuum and while being diffracted within crystals (Werner, 1980) (see Section 5.3.5). Experiments completely different from the X-ray case can thus be performed with perfect crystals and with neutron interferometers (see Sections 5.3.6 and 5.3.7.3).

5.3.2.5. Translating X-ray dynamical theory into the neutron case

As shown in Chapter 5.1, the basic equations of dynamical theory, *viz* Maxwell's equations for the X-ray case and the time-independent Schrödinger equation in the neutron case, have exactly the same form when the effect of the neutron spin can be neglected, *i.e.* in situations that do not involve magnetism and when no externally applied potential is taken into account. The translation scheme for the scattering factors and structure factors is described above. The one formal difference is that the wavefunction is scalar in the neutron case, hence there is no equivalent to the parallel and perpendicular polarizations of the X-ray situation: C in equation (5.1.2.20) of Chapter 5.1 should therefore be set to 1.

The physics of neutron diffraction by perfect crystals is therefore expected to be very similar to that of X-ray diffraction, with the existence of wavefields, *Pendellösung* effects, anomalous transmission, intrinsic rocking-curve shapes and reflectivity *versus* thickness behaviour in direct correspondence. All experimental tests of these predictions confirm this view (Section 5.3.6).

Basic discussions of dynamical neutron scattering are given by Stassis & Oberteuffer (1974), Sears (1978), Rauch & Petrascheck (1978) and Squires (1978).

5.3.3. Neutron spin, and diffraction by perfect magnetic crystals

5.3.3.1. Polarization of a neutron beam and the Larmor precession in a uniform magnetic field

A polarized neutron beam is represented by a two-component spinor,

$$|\varphi\rangle = \begin{pmatrix} c \\ d \end{pmatrix} = c\begin{pmatrix} 1 \\ 0 \end{pmatrix} + d\begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

which is the coherent superposition of two states, of different amplitudes c and d, polarized in opposite directions along the spin-quantization axis. The spinor components c and d are generally space- and time-dependent. We suppose that $\langle\varphi|\varphi\rangle = cc^* + dd^* = 1$. The polarization vector $\mathbf{P}$ is defined as

$$\mathbf{P} = \langle\varphi|\boldsymbol{\sigma}|\varphi\rangle,$$

where the vector $\boldsymbol{\sigma}$ represents the set of Pauli matrices σ_x, σ_y and σ_z. The components of $\mathbf{P}$ are

$$\begin{aligned} P_x &= \begin{pmatrix} c^* & d^* \end{pmatrix}\sigma_x \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} c^* & d^* \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} \\ &= c^*d + cd^* \\ P_y &= \begin{pmatrix} c^* & d^* \end{pmatrix}\sigma_y \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} c^* & d^* \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} \\ &= i(cd^* - c^*d) \\ P_z &= \begin{pmatrix} c^* & d^* \end{pmatrix}\sigma_z \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} c^* & d^* \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} \\ &= cc^* - dd^*, \end{aligned} \tag{5.3.3.1}$$

from which it is clearly seen that, unlike P_z, the polarization components P_x and P_y depend on the phase difference between the spinor components c and d.

In a region of a vacuum in which a uniform magnetic field $\mathbf{B}$ is present, a neutron beam experiences a magnetic potential energy represented by the matrix

$$-\mu_n \boldsymbol{\sigma}\cdot\mathbf{B} = \begin{pmatrix} -\mu_n B & 0 \\ 0 & \mu_n B \end{pmatrix},$$

μ_n being the neutron magnetic moment, if the directions of $\mathbf{B}$ and of the spin-quantization axis coincide. Consequently, different indices of refraction $n = 1 \pm (\mu_n B/2E)$, where E is the neutron energy, should be associated with the spinor components c and d; this induces between these spinor components a phase difference which is a linear function of the time (or, equivalently, of the distance travelled by the neutrons), hence, according to (5.3.3.1), a rotation around the magnetic field of the component of the neutron polarization perpendicular to this magnetic field. The time frequency of this so-called Larmor precession is $2\mu_n B/h$, where h is Planck's constant.

A neutron beam may be partially polarized; such a beam is conveniently represented by a spin-density matrix ρ, which is the statistical average of the spin-density matrices associated to the polarized beams which are mixed incoherently, the density matrix associated to the spinor

$$|\varphi\rangle = \begin{pmatrix} c \\ d \end{pmatrix}$$

being

$$|\varphi\rangle\langle\varphi| = \begin{pmatrix} c \\ d \end{pmatrix}\begin{pmatrix} c^* & d^* \end{pmatrix} = \begin{pmatrix} cc^* & cd^* \\ c^*d & dd^* \end{pmatrix}.$$

The polarization vector $\mathbf{P}$ is then obtained as

$$\mathbf{P} = \mathrm{Tr}(\boldsymbol{\sigma}\rho). \tag{5.3.3.2}$$

In the common case of a nonpolarized beam, the spin-density matrix is

$$\rho = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is easily seen that all components of $\mathbf{P}$ are then equal to 0.

Equation (5.3.3.2) is therefore applicable to the general case (polarized, partially polarized or nonpolarized beam). The inverse relation giving the density matrix ρ as function of $\mathbf{P}$ is

$$\rho = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \tfrac{1}{2}\mathbf{P}\cdot\boldsymbol{\sigma}. \tag{5.3.3.3}$$

5.3.3.2. Magnetic scattering by a single ion having unpaired electrons

The spin and orbital motion of unpaired electrons in an atom or ion give rise to a surrounding magnetic field $\mathbf{B}(\mathbf{r})$ which acts on the neutron *via* the magnetic potential energy $-\boldsymbol{\mu}_n\cdot\mathbf{B}(\mathbf{r})$, where $\boldsymbol{\mu}_n$ is the neutron magnetic moment. Since this is a long-range interaction, in contrast to the nuclear interaction, the magnetic scattering length p, which is proportional to the Fourier transform of the magnetic potential energy distribution $-\boldsymbol{\mu}_n\cdot\mathbf{B}(\mathbf{r})$, depends on the angle of scattering.

The classical relation $\mathrm{div}\ \mathbf{B}(\mathbf{r}) = 0$ shows clearly that the vector $\mathbf{B}(\mathbf{s})$, which is the Fourier transform of $\mathbf{B}(\mathbf{r})$, is perpendicular to the reciprocal-space vector $\mathbf{s}$. If we consider the magnetic field $\mathbf{B}(\mathbf{r})$ as resulting from a point-like magnetic moment $\boldsymbol{\mu}$ at position $\mathbf{r} = 0$, we get

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\mathrm{curl}\frac{\boldsymbol{\mu}\times\mathbf{r}}{r^3},$$

where $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1} is the permittivity of a vacuum and $\times$ denotes the cross product. $\mathbf{B}(\mathbf{r})$ can be Fourier-transformed into

$$\mathbf{B}(\mathbf{s}) = \mu_0\mathbf{s}\times\frac{\boldsymbol{\mu}\times\mathbf{s}}{s^2} = \mu_0\boldsymbol{\mu}_\perp(\mathbf{s}), \tag{5.3.3.4}$$

where $\boldsymbol{\mu}_\perp(\mathbf{s})$ is the projection of $\boldsymbol{\mu}$ on the plane perpendicular to $\mathbf{s}$ (reflecting plane).

This result can be applied by volume integration to the more general case of a spatially extended magnetization distribution, which for a single magnetic ion corresponds to the atomic shell of the unpaired electrons. It is thus shown that the magnetic scattering length is proportional to $\boldsymbol{\mu}_n\cdot\boldsymbol{\mu}_{i\perp}$, where $\boldsymbol{\mu}_{i\perp}$ is the projection of the magnetic moment of the ion on the reflecting plane.

For a complete description of magnetic scattering, which involves the spin-polarization properties of the scattered beam, it is necessary to represent the neutron wavefunction in the form of a two-component spinor and the ion's magnetic moment as a spin operator which is a matrix expressed in terms of the Pauli matrices $\boldsymbol{\sigma}$ $(\sigma_x, \sigma_y, \sigma_z)$. The magnetic scattering length is therefore itself a (2×2) matrix:

$$(\mathrm{p}) = -(2\pi m/h^2)\mu_n\boldsymbol{\sigma}\cdot\mathbf{B}(\mathbf{s}) = -\mu_0(2\pi m/h^2)\mu_n\boldsymbol{\sigma}\cdot\boldsymbol{\mu}_{i\perp}(\mathbf{s})f_i(\sin\theta/\lambda), \tag{5.3.3.5}$$

where $f_i(\sin\theta/\lambda)$ is the dimensionless magnetic form factor of the ion considered and tends towards a maximum value of 1 when the scattering angle θ tends towards 0 (forward scattering). The value of $\mu_0(2\pi m/h^2)\mu_n\mu_i$ is $p_1 = 2.70 \times 10^{-15}$ m for $\mu_i = 1$ Bohr magneton.

According to (5.3.3.4) or (5.3.3.5), there is no magnetic scattering in directions such that the scattering vector $\mathbf{s}$ is in the same direction as the ion magnetic moment $\boldsymbol{\mu}_i$. Magnetic scattering effects are maximum when $\mathbf{s}$ and $\boldsymbol{\mu}_i$ are perpendicular.

The matrix (p) is diagonal if the direction of $\boldsymbol{\mu}_{i\perp}(\mathbf{s})$ is chosen as the spin-quantization axis. Therefore, there is no spin-flip scattering if the incident beam is polarized parallel or antiparallel to the direction of $\boldsymbol{\mu}_{i\perp}(\mathbf{s})$.

It is more usual to choose the spin-quantization axis (Oz) along $\boldsymbol{\mu}_i$. Let β be the angle between the vectors $\boldsymbol{\mu}_i$ and $\mathbf{s}$; the

(x, y, z) components of $\boldsymbol{\mu}_{i\perp}(\mathbf{s})$ are then $(-\mu_i \sin\beta\cos\beta,\ 0,\ \mu_i \sin^2\beta)$ if the y axis is chosen along $\boldsymbol{\mu}_i \times \mathbf{s}$. The total scattering length, which is the sum of the nuclear and the magnetic scattering lengths, is then represented by the matrix

$$(\mathrm{q}) = \begin{pmatrix} b + p\sin^2\beta & -p\sin\beta\cos\beta \\ -p\sin\beta\cos\beta & b - p\sin^2\beta \end{pmatrix}, \qquad (5.3.3.6)$$

where b is the nuclear scattering length and

$$p = -\mu_0 \frac{2\pi m}{h^2} \mu_n \mu_i f_i\left(\frac{\sin\theta}{\lambda}\right) = -p_1 \mu_i f_i\left(\frac{\sin\theta}{\lambda}\right),$$

with μ_i expressed in Bohr magnetons. The relations

$$(\mathrm{q})\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix} b + p\sin^2\beta \\ -p\sin\beta\cos\beta \end{pmatrix} \text{ and}$$
$$(\mathrm{q})\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix} -p\sin\beta\cos\beta \\ b - p\sin^2\beta \end{pmatrix}$$

show clearly that the diagonal and the nondiagonal elements of the matrix (q) are, respectively, the spin-flip and the non-spin-flip scattering lengths. It is usual to consider the scattering cross sections, which are the measurable quantities. The cross sections for neutrons polarized parallel or antiparallel to the ion magnetic moment are

$$(\mathrm{d}\sigma/\mathrm{d}\Omega)_{\pm} = b^2 \pm 2bp\sin^2\beta + (p\sin\beta)^2. \qquad (5.3.3.7)$$

These expressions are the sum of the spin-flip and non-spin-flip cross sections, which are equal to $(b \pm p\sin^2\beta)^2$ and $(p\sin\beta\cos\beta)^2$, respectively. In the case of nonpolarized neutrons, the interference term $(\pm 2bp\sin^2\beta)$ between the nuclear and the magnetic scattering disappears; the cross section is then

$$(\mathrm{d}\sigma/\mathrm{d}\Omega) = b^2 + (p\sin\beta)^2. \qquad (5.3.3.8)$$

In the general case of a partially polarized beam we can use the density-matrix representation. Let ρ_{inc} be the density matrix of the incident beam; it can be shown that the density matrix of the diffracted beam is equal to the following product of matrices: $(\mathrm{q})\rho_{inc}(\mathrm{q}^*)$. Using the relations between the density matrix and polarization vector presented in the preceding section, we can obtain a general description of the diffracted beam as a function of the polarization properties of the incident beam. Such a formalism is of interest for dealing with new experimental arrangements, in which a three-dimensional polarization analysis of the diffracted beam is possible, as shown by Tasset (1989).

5.3.3.3. Dynamical theory in the case of perfect ferromagnetic or collinear ferrimagnetic crystals

The most direct way to develop this dynamical theory in the two-beam case, which involves a single Bragg-diffracted beam of diffraction vector $\mathbf{h}$, is to consider spinor wavefunctions of the following form:

$$\varphi(\mathbf{r}) = \exp(i\mathbf{K_0}\cdot\mathbf{r})\begin{pmatrix}D_0\\E_0\end{pmatrix} + \exp[i(\mathbf{K_0}+\mathbf{h})\mathbf{r}]\begin{pmatrix}D_h\\E_h\end{pmatrix} \qquad (5.3.3.9)$$

as approximate solutions of the wave equation inside the crystal,

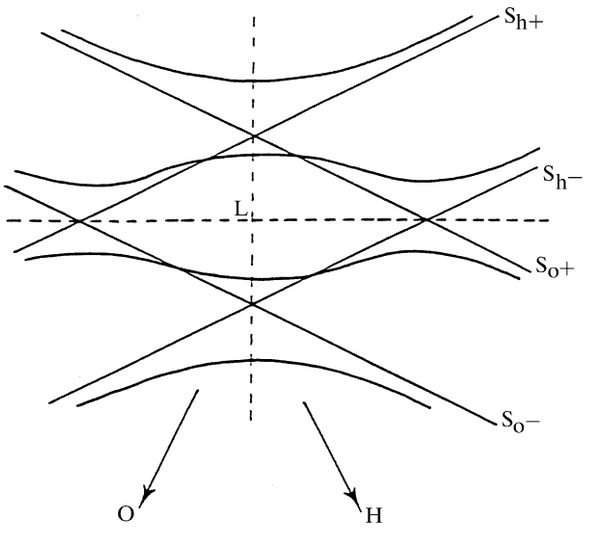


Fig. 5.3.3.1. Schematic plot of the two-beam dispersion surface in the case of a purely magnetic reflection such that $\mathbf{Q_h} = \mathbf{Q_{-h}} = \mathbf{Q_0}$ and that the angle between $\mathbf{Q_0}$ and $\mathbf{Q_h}$ is equal to $\pi/4$.

$$\Delta\varphi(\mathbf{r}) + k^2\varphi(\mathbf{r}) = [u(\mathbf{r}) - \boldsymbol{\sigma}\cdot\mathbf{Q}(\mathbf{r})]\varphi(\mathbf{r}), \qquad (5.3.3.10)$$

where $u(\mathbf{r})$ and $-\boldsymbol{\sigma}\cdot\mathbf{Q}(\mathbf{r})$ are, respectively, equal to the nuclear and the magnetic potential energies multiplied by $2m/\hbar^2$. In the calculation of $\varphi(\mathbf{r})$ in the two-beam case, we need only three terms in the expansions of the functions $u(\mathbf{r})$ and $\mathbf{Q}(\mathbf{r})$ into Fourier series:

$$u(\mathbf{r}) = u_0 + u_{\mathbf{h}}\exp(i\mathbf{h}\cdot\mathbf{r}) + u_{-\mathbf{h}}\exp(-i\mathbf{h}\cdot\mathbf{r}) + \ldots,$$
$$\mathbf{Q}(\mathbf{r}) = \mathbf{Q_0} + \mathbf{Q_h}\exp(i\mathbf{h}\cdot\mathbf{r}) + \mathbf{Q_{-h}}\exp(-i\mathbf{h}\cdot\mathbf{r}) + \ldots.$$

We suppose that the crystal is magnetically saturated by an externally applied magnetic field $\mathbf{H}_a$. $\mathbf{Q_0}$ is then proportional to the macroscopic mean magnetic field $\mathbf{B} = \mu_0(\mathbf{M} + \mathbf{H}_a + \mathbf{H}_d)$, where $\mathbf{M}$ is the magnetization vector and $\mathbf{H}_d$ is the demagnetizing field. The results of Section 5.3.3.2 show that $\mathbf{Q_h}$ and $\mathbf{Q_{-h}}$ are proportional to the projection of $\mathbf{M}$ on the reflecting plane.

The four coefficients D_0, D_h, E_0 and E_h of (5.3.3.9) are found to satisfy a system of four homogeneous linear equations. The condition that the associated determinant has to be equal to 0 defines the dispersion surface, which is of order 4 and has four branches. An incident plane wave thus excites a system of four wavefields of the form of (5.3.3.9), generally polarized in various directions. A particular example of a dispersion surface, having an unusual shape, is shown in Fig. 5.3.3.1.

This is a much more complicated situation than in the case of nonmagnetic crystals, in which one only needs to consider scalar wavefunctions which depend on two coefficients, such as D_0 and D_h, and which are related to hyperbolic dispersion surfaces of order 2, as fully described in Chapter 5.1 on X-ray diffraction.

In fact, all neutron experiments related to dynamical effects in diffraction by magnetic crystals have been performed under such conditions that the magnetization vector in the crystal is perpendicular to the diffraction vector $\mathbf{h}$. In this case, the vectors $\mathbf{Q_h}$ and $\mathbf{Q_{-h}}$ are parallel or antiparallel to the vector $\mathbf{Q_0}$ which is chosen as the spin-quantization axis. The matrices $\boldsymbol{\sigma}\cdot\mathbf{Q_0}$, $\boldsymbol{\sigma}\cdot\mathbf{Q_h}$ and $\boldsymbol{\sigma}\cdot\mathbf{Q_{-h}}$ are then all diagonal matrices, and we obtain for the two spin states $(\pm)$ separate dynamical equations which are similar to the dynamical equations for the scalar case, but with different structure factors, which are either the sum or the difference of the nuclear structure factor F_N and of the magnetic structure factor F_M:

$$F_+ = F_N + F_M \text{ and } F_- = F_N - F_M. \quad (5.3.3.11)$$

F_N and F_M are related to the scattering lengths of the ions in the unit cell of volume V_c:

$$F_N = V_c u_{\mathbf{h}} = \sum_i b_i \exp(-i\mathbf{h}\cdot\mathbf{r}_i);$$

$$F_M = V_c|\mathbf{Q}_{\mathbf{h}}| = -\frac{\mu_0 m}{2h^2}\mu_n \boldsymbol{\sigma}\cdot\sum_i \boldsymbol{\mu}_{i\perp}(\mathbf{h}) f_i\left(\frac{\sin\theta}{\lambda}\right)\exp(-i\mathbf{h}\cdot\mathbf{r}_i).$$

The dispersion surface of order 4 degenerates into two hyperbolic dispersion surfaces, each of them corresponding to one of the polarization states (±). The asymptotes are different; this is related to different values of the refractive indices for neutron polarization parallel or antiparallel to $\mathbf{Q}_0$.

In some special cases the magnitudes of F_N and F_M happen to be equal. Only one polarization state is then reflected. Magnetic crystals with such a property (reflections 111 of the Heusler alloy Cu_2MnAl, or 200 of the alloy Co–8% Fe) are very useful as polarizing monochromators and as analysers of polarization.

If the scattering vector $\mathbf{h}$ is in the same direction as the magnetization, this reflection is a purely nuclear one (with no magnetic contribution), since F_M is then equal to 0. Purely magnetic reflections (without nuclear contribution) also exist if the magnetic structure involves several sublattices.

If $\mathbf{h}$ is neither perpendicular to the average magnetization nor in the same direction, the presence of nondiagonal matrices in the dynamical equations cannot be avoided. The dynamical theory of diffraction by perfect magnetic crystals then takes the complicated form already mentioned.

Theoretical discussions of this complicated case of dynamical diffraction have been given by Stassis & Oberteuffer (1974), Mendiratta & Blume (1976), Sivardière (1975), Belyakov & Bokun (1975, 1976), Schmidt *et al.* (1975), Bokun (1979), Guigay & Schlenker (1979*a*,*b*) and Schmidt (1983). However, to our knowledge, only limited experimental work has been carried out on this subject. Successful experiments could only be performed for the simpler cases mentioned above.

5.3.3.4. The dynamical theory in the case of perfect collinear antiferromagnetic crystals

In this case, there is no average magnetization ($\mathbf{Q}_0 = 0$). It is then convenient to choose the quantization axis in the direction of $\mathbf{Q}_{\mathbf{h}}$ and $\mathbf{Q}_{-\mathbf{h}}$. The dispersion surface degenerates into two hyperbolic surfaces corresponding to each polarization state along this direction for any orientation of the diffraction vector relative to the direction of the magnetic moments of the sublattices. These two hyperbolic dispersion surfaces have the same asymptotes. Furthermore, in the case of a purely magnetic reflection, they are identical.

The possibility of observing a precession of the neutron polarization in the presence of diffraction, in spite of the fact that there is no average magnetization, has been pointed out by Baryshevskii (1976).

5.3.3.5. The flipping ratio

In polarized neutron diffraction by a magnetically saturated magnetic sample, it is usual to measure the ratio of the reflected intensities I_+ and I_- measured when the incident beam is polarized parallel or antiparallel to the magnetization in the sample. This ratio is called the flipping ratio,

$$R = I_+/I_-, \quad (5.3.3.12)$$

because its measurement involves flipping the incident-beam polarization to the opposite direction. This is an experimentally well defined quantity, because it is independent of a number of parameters such as the intensity of the incident beam, the temperature factor or the coefficient of absorption. In the case of an ideally imperfect crystal, we obtain from the kinematical expressions of the integrated reflectivities

$$R_{\text{kin}}(\mathbf{h}) = (I_+/I_-)_{\text{kin}} = \left(\frac{|F_N + F_M|}{|F_N - F_M|}\right)^2. \quad (5.3.3.13)$$

In the case of an ideally perfect thick crystal, we obtain from the dynamical expressions of the integrated reflectivities

$$R_{\text{dyn}}(\mathbf{h}) = (I_+/I_-)_{\text{dyn}} = \frac{|F_N + F_M|}{|F_N - F_M|}. \quad (5.3.3.14)$$

In general, R_{dyn} depends on the wavelength and on the crystal thickness; these dependences disappear, as seen from (5.3.3.14), if the path length in the crystal is much larger than the extinction distances for the two polarization states. It is clear that the determination of R_{kin} or R_{dyn} allows the determination of the ratio F_M/F_N, hence of F_M if F_N is known. In fact, because real crystals are neither ideally imperfect nor ideally perfect, one usually introduces an extinction factor y (extinction is discussed below, in Section 5.3.4) in order to distinguish the real crystal reflectivity from the reflectivity of the ideally imperfect crystal. Different extinction coefficients y_+ and y_- are actually expected for the two polarization states. This obviously complicates the task of the determination of F_M/F_N.

In the kinematical approximation, the flipping ratio does not depend on the wavelength, in contrast to dynamical calculations for hypothetically perfect crystals (especially for the Laue case of diffraction). Therefore, an experimental investigation of the wavelength dependence of the flipping ratio is a convenient test for the presence of extinction. Measurements of the flipping ratio have been used by Bonnet *et al.* (1976) and by Kulda *et al.* (1991) in order to test extinction models. Baruchel *et al.* (1986) have compared nuclear and magnetic extinction in a crystal of MnP.

Instead of considering only the ratio of the integrated reflectivities, it is also possible to record the flipping ratio as a function of the angular position of the crystal as it is rotated across the Bragg position. Extinction is expected to be maximum at the peak and the ratio measured on the tails of the rocking curve may approach the kinematical value. It has been found experimentally that this expectation is not of general validity, as discussed by Chakravarthy & Madhav Rao (1980). It would be valid in the case of a perfect crystal, hence in the case of pure primary extinction. It would also be valid in the case of secondary extinction of type I, but not in the case of secondary extinction of type II [following Zachariasen (1967), type II corresponds to mosaic crystals such that the diffraction pattern from each block is wider than the mosaic statistical distribution].

5.3.4. Extinction in neutron diffraction (nonmagnetic case)

The kinematical approximation, which corresponds to the first Born approximation in scattering theory, supposes that each incident neutron can be scattered only once and therefore neglects the possibility that the neutrons may be scattered several times. Because this is a simple approximation which overestimates the crystal reflectivity, the actual reduction of reflectivity, as compared to its kinematical value, is termed *extinction.* This is actually a typical dynamical effect, since it is a multiple-scattering effect.

Extinction effects can be safely neglected in the case of scattering by very small crystals; more precisely, this is possible when the path length of the neutron beam in the crystal is much smaller than $\Delta = V_c/\lambda F$, where λ is the neutron wavelength and F/V_c is the scattering length per unit volume for the reflection considered. Δ is sometimes called the 'extinction distance'.

A very important fact is that extinction effects also vanish if the crystal is imperfect enough, because each plane-wave component of the incident beam can then be Bragg-reflected in only a small volume of the sample. This is the extinction-free case of 'ideally imperfect crystals'. Conversely, extinction is maximum (smallest value of y) in the case of ideally perfect non-absorbing crystals.

Clearly, no significant extinction effects are expected if the crystal is thick but strongly absorbing, more precisely if the linear absorption coefficient μ is such that $\mu\Delta \gg 1$. Neutron diffraction usually corresponds to the opposite case ($\mu\Delta \ll 1$), in which extinction effects in nearly perfect crystals dominate absorption effects.

Extinction effects are usually described in the frame of the mosaic model, in which the crystal is considered as a juxtaposition of perfect blocks with different orientations. The relevance of this model to the case of neutron diffraction was first considered by Bacon & Lowde (1948). If the mosaic blocks are big enough there is extinction within each block; this is called *primary extinction.* Multiple scattering can also occur in different blocks if their misorientation is small enough. In this case, which is called *secondary extinction*, there is no phase coherence between the scattering events in the different blocks. The fact that empirical intensity-coupling equations are used in this case is based on this phase incoherence.

In the general case, primary and secondary extinction effects coexist. Pure secondary extinction occurs in the case of a mosaic crystal made of very small blocks. Pure primary extinction is observed in diffraction by perfect crystals.

The parameters of the mosaic model are the average size of the perfect blocks and the angular width of their misorientation distribution. The extinction theory of the mosaic model provides a relation between these parameters and the extinction coefficient, defined as the ratio of the observed reflectivity to the ideal one, which is the kinematical reflectivity in this context.

In conventional work, the crystal structure factors of different reflections and the parameters of the mosaic model are fitted together to the experimental data, which are the integrated reflectivities and the angular widths of the rocking curves. In many cases, only the weakest reflections will be free, or nearly free, from extinction. The extinction corrections thus obtained can be considered as satisfactory in cases of moderate extinction. Nevertheless, extinction remains a real problem in cases of strong extinction and in any case if a very precise determination of the crystal structure factors is required.

There exist several forms of the mosaic model of extinction. For instance, in the model developed by Kulda (1988*a*,*b*, 1991), the mosaic blocks are not considered just as simple perfect blocks but may be deformed perfect blocks. This has the advantage of including the case of macroscopically deformed crystals, such as bent crystals.

A basically different approach, free from the distinction between primary and secondary extinction, has been proposed by Kato (1980*a*,*b*). This is a wave-optical approach starting from the dynamical equations for diffraction by deformed crystals. These so-called Takagi–Taupin equations (Takagi, 1962; Taupin, 1964) contain a position-dependent phase factor related to the displacement field of the deformed crystal lattice. Kato proposed considering this phase factor as a random function with suitably defined statistical characteristics. The wave amplitudes are then also random functions, the average of which represent the coherent wavefields while their statistical fluctuations represent the incoherent intensity fields.

Modifications to the Kato formulation have been introduced by Al Haddad & Becker (1988), by Becker & Al Haddad (1990, 1992), by Guigay (1989) and by Guigay & Chukhovskii (1992, 1995). Presently, it is not easy to apply this 'statistical dynamical theory' to real experiments. The widely used methods for extinction corrections are still based on the former mosaic model, according to the formulation of Zachariasen (1967), later improved by Becker & Coppens (1974*a*,*b*, 1975).

As in the X-ray case, acoustic waves produced by ultrasonic excitation can artificially induce a transition from perfect to ideally imperfect crystal behaviour. The effect of ultrasound on the scattering behaviour of distorted crystals is quite complex. A good discussion with reference to neutron-scattering experiments is given by Zolotoyabko & Sander (1995).

The situation of crystals with a simple distortion field is less difficult than the statistical problem of extinction. Klar & Rustichelli (1973) confirmed that the Takagi–Taupin equations, originally devised for X-rays, can be used for neutron diffraction with due account of the very small absorption, and used them for computing the effect of crystal curvature.

5.3.5. Effect of external fields on neutron scattering by perfect crystals

The possibility of acting on neutrons through externally applied fields during their propagation in perfect crystals provides possibilities that are totally unknown in the X-ray case. The theory has been given by Werner (1980) using the approaches (migration of tie points, and Takagi–Taupin equations) that are customary in the treatment of imperfect crystals (see above). Zeilinger *et al.* (1986) pointed out that the effective-mass concept, familiar in describing electrons in solid-state physics, can shed new light on this behaviour: because of the curvature of the dispersion surface at a near-exact Bragg setting, effective masses five orders of magnitude smaller than the rest mass of the neutron in a vacuum can be obtained. Related experiments are discussed below.

An interesting proposal was put forward by Horne *et al.* (1988) on the coupling between the Larmor precession in a homogeneous magnetic field and the spin–orbit interaction of the neutron with nonmagnetic atoms, a term which was dismissed in Section 5.3.2 because its contribution to the scattering length is two orders of magnitude smaller than that of the nuclear term. A resonance is expected to show up as highly enhanced diffracted intensity when a perfect sample is set for Bragg scattering and the magnetic field is adjusted so that the Larmor precession period is equal to the *Pendellösung* period.

5.3.6. Experimental tests of the dynamical theory of neutron scattering

These experiments are less extensive for neutron scattering than for X-rays. The two most striking effects of dynamical theory for nonmagnetic nearly perfect crystals, *Pendellösung* behaviour and anomalous absorption, have been demonstrated in the neutron case too. *Pendellösung* measurement is described below (Section 5.3.7.2) because it is useful in the determination of scattering lengths. The anomalous transmission effect occurring when a perfect absorbing crystal is exactly at Bragg setting, *i.e.* the Borrmann effect, is often referred to in the neutron case as the suppression of the inelastic channel in resonance scattering, after Kagan & Afanas'ev (1966), who worked out the theory. A small decrease in absorption was detected in pioneering experiments on calcite by Knowles (1956) using the corresponding decrease in the emission of γ-rays and by Sippel *et al.* (1962), Shil'shtein *et al.* (1971) and Hastings *et al.* (1990) directly. Rocking curves of perfect crystals were measured by Sippel *et al.* (1964) in transmission, and by Kikuta *et al.* (1975). Integrated intensities were

measured by Lambert & Malgrange (1982). The large angular amplification associated with the curvature of the dispersion surfaces near the exact Bragg setting was demonstrated by Kikuta *et al.* (1975) and by Zeilinger & Shull (1979).

In magnetic crystals, the investigations have been restricted to the simpler geometry where the scattering vector is perpendicular to the magnetization, and to few materials. *Pendellösung* behaviour was evidenced through the variation with wavelength of the flipping ratio for polarized neutrons by Baruchel *et al.* (1982) on an yttrium iron garnet sample, with the geometry selected so that the defects would not affect the Bragg reflection used. The inclination method was used successfully by Zelepukhin *et al.* (1989), Kvardakov & Somenkov (1990) and Kvardakov *et al.* (1990*a*) for the weak ferromagnet $FeBO_3$, and in the room-temperature weak-ferromagnetic phase of hematite, α-Fe_2O_3, by Kvardakov *et al.* (1990*b*) and Kvardakov & Somenkov (1992).

Experiments on the influence of defects in nearly perfect crystals have been performed by several groups. The effect on the rocking curve was investigated by Eichhorn *et al.* (1967), the intensities were measured by Lambert & Malgrange (1982) and by Albertini, Boeuf, Cesini *et al.* (1976), and the influence on the *Pendellösung* behaviour was discussed by Kvardakov & Somenkov (1992). Boeuf & Rustichelli (1974) and Albertini *et al.* (1977) investigated silicon crystals curved by a thin surface silicon nitride layer.

Many experiments have been performed on vibrating crystals; reviews are given by Michalec *et al.* (1988) and by Kulda *et al.* (1988). Because the velocity of neutrons is of the same order of magnitude as the velocity of acoustic phonons in crystals, the effect of ultrasonic excitation on dynamical diffraction takes on some original features compared to the X-ray case (Iolin & Entin, 1983); they could to some extent be evidenced experimentally (Iolin *et al.*, 1986; Chalupa *et al.*, 1986). References to experimental work on neutron scattering by imperfect crystals under ultrasonic excitation are included in Zolotoyabko & Sander (1995).

Some experiments with no equivalent in the X-ray case could be performed. The very strong incoherent scattering of neutrons by protons, very different physically but similar in its effect to absorption, was also shown to lead to anomalous transmission effects by Sippel & Eichhorn (1968). Because the velocity of thermal neutrons in a vacuum is five orders of magnitude smaller than the velocity of light, the flight time for neutrons undergoing Bragg scattering in Laue geometry in a perfect crystal could be measured directly (Shull *et al.*, 1980). The effect of externally applied fields was measured experimentally for magnetic fields by Zeilinger & Shull (1979) and Zeilinger *et al.* (1986). Slight rotation of the crystal, introducing a Coriolis force, was used by Raum *et al.* (1995), and gravity was tested recently, with the spectacular result that some states are accelerated upwards (Zeilinger, 1995).

5.3.7. Applications of the dynamical theory of neutron scattering

5.3.7.1. Neutron optics

Most experiments in neutron scattering require an intensity-effective use of the available beam at the cost of relatively high divergence and wavelength spread. The monochromators must then be imperfect ('mosaic') crystals. In some cases, however, it is important to have a small divergence and wavelength band. One example is the search for small variations in neutron energy in inelastic scattering without the use of the neutron spin-echo principle. Perfect crystals must then be used as monochromators or analysers, and dynamical diffraction is directly involved. As in the X-ray case, special designs can lead to strong decrease in the intensity of harmonics, *i.e.* of contributions of $\lambda/2$ or $\lambda/3$ (Hart & Rodrigues, 1978). The possibility of focusing neutron beams by the use of perfect crystals with the incident beam spatially modulated in amplitude through an absorber, or in phase through an appropriate patterning of the surface, in analogy with the Bragg–Fresnel lenses developed for X-rays, was suggested by Indenbom (1979).

The use of two identical perfect crystals in nondispersive (+, −, ‖) setting provides a way of measuring the very narrow intrinsic rocking curves expected from the dynamical theory. Any divergence added between the two crystals can be sensitively measured. Thus perfect crystals provide interesting possibilities for measuring very small angle neutron scattering. This was performed by Takahashi *et al.* (1981, 1983) and Tomimitsu *et al.* (1986) on amorphous materials, and by Kvardakov *et al.* (1987) for the investigation of ferromagnetic domains in bulk silicon–iron specimens under stress, both through the variations in transmission associated with refraction on the domain walls and through small-angle scattering. Imaging applications are described in Section 5.3.7.4. Badurek *et al.* (1979) used the different deflection of the two polarization states provided by a magnetic prism placed between two perfect silicon crystals to produce polarized beams.

Curved almost-perfect crystals or crystals with a gradient in the lattice spacing can provide focusing (Albertini, Boeuf, Lagomarsino *et al.*, 1976) and vibrating crystals can give the possibility of tailoring the reflectivity of crystals, as well as of modulating beams in time (Michalec *et al.*, 1988). A double-crystal arrangement with bent crystals was shown by Eichhorn (1988) to be a flexible small-angle-neutron-scattering device.

5.3.7.2. Measurement of scattering lengths by Pendellösung effects

As in X-ray diffraction, *Pendellösung* oscillations provide an accurate way of measuring structure factors, hence coherent neutron scattering lengths. The equal-thickness fringes expected from a wedge-shaped crystal were observed by Kikuta *et al.* (1971). Three kinds of measurements were made. Sippel *et al.* (1965) measured as a function of thickness the integrated reflectivity from a perfect crystal of silicon, the thickness of which they varied by polishing after each measurement, obtaining a curve similar to Fig. 5.1.6.7, corresponding to equation 5.1.6.8. Shull (1968) restricted the measurement to wavefields that propagated along the reflecting planes, hence at exact Bragg incidence, by setting fine slits on the entrance and exit faces of 3 to 10 mm-thick silicon crystals, and measured the oscillation in diffracted intensity as he varied the wavelength of the neutrons used by rotating the crystal. Shull & Oberteuffer (1972) showed that a better interpretation of the data, when the beam is restricted to a fine slit, corresponds to the spherical wave approach (actually cylindrical wave), and the boundary conditions were discussed more generally by Arthur & Horne (1985). Somenkov *et al.* (1978) developed the inclination method, in which the integrated reflectivity is measured as the effective crystal thickness is varied nondestructively, by rotating the crystal around the diffraction vector, and used it for germanium. Belova *et al.* (1983) discuss this method in detail. The results obtained by this group for magnetic crystals are mentioned in Section 5.3.6. Structure-factor values for magnetic reflections were obtained by Kvardakov *et al.* (1995) for the weak ferromagnet $FeBO_3$.

5.3.7.3. Neutron interferometry

Because diffraction by perfect crystals provides a well defined distribution of the intensity and phase of the beam, interferometry with X-rays or neutrons is possible using ingeniously designed and carefully manufactured monolithic devices carved out of single crystals of silicon. The technical and scientific features of this family of techniques are well summarized by Bonse (1979, 1988), as well as other papers in the same volumes, and by Shull (1986).

X-ray interferometry started with the Bonse–Hart interferometer (Bonse & Hart, 1965). A typical device is the LLL skew-symmetric interferometer, where the L's stand for Laue,

indicating transmission geometry in all crystal slabs. In these slabs, which can be called the splitter, the mirrors and the recombiner, the same pair of opposite reflections, in symmetrical Laue geometry, is used three times. In the first slab, the incident beam is coherently split into a transmitted and a diffracted beam. Each of these is then diffracted in the two mirrors, and the resulting beams interfere in the recombiner, again yielding a forward-diffracted and a diffracted beam, the intensities of both of which are measured. This version, the analogue of the Mach–Zehnder interferometer in optics, offers a sizeable space (several cm of path length) where two coherent parallel beams can be submitted to various external actions. Shifting the relative phase of these beams (*e.g.* by π, introducing an optical path-length difference of $\lambda/2$) results in the intensities of the outgoing beams changing from a maximum to a minimum.

Applications of neutron interferometry range from the very useful to the very exotic. The most useful one is probably the measurement of coherent neutron scattering lengths. Unlike the *Pendellösung* method described in Section 5.3.7.2, this method does not require the measured samples to be perfect single crystals, nor indeed crystals. Placing a slab of material across one of the beams and rotating it will induce an optical path-length difference of $(1 - n)t$ if t is the effective thickness along the beam, hence a phase shift of $2\pi(1 - n)t/\lambda$. With the expression of the refractive index n as given in Section 5.3.2.2, it is clear that for an isotopically pure material the scattering length $b_{\rm coh}$ can be deduced from the measurement of intensity *versus* the rotation angle of the phase shifter. This is a very versatile and much used method. The decrease in oscillation contrast can be used to obtain information of relevance to materials science, such as statistical properties of magnetic domain distributions (Korpiun, 1966) or precipitates (Rauch & Seidl, 1987); Rauch (1995) analyses the effect in terms of the neutron coherence function.

Many elegant experiments have been performed with neutron interferometers in efforts to set an upper limit to effects than can be considered as nonexistent, or to test expectations of basic quantum physics. Many papers are found in the same volumes as Bonse (1979) and Bonse (1988); excellent reviews have been given by Klein & Werner (1983), Klein (1988) and Werner (1995). Among the topics investigated are the effect of gravity (Colella *et al.*, 1975), the Sagnac effect, *i.e.* the influence of the Earth's rotation (Werner *et al.*, 1979), the Fizeau effect, *i.e.* the effect of the movement of the material through which the neutrons are transmitted (Arif *et al.*, 1988) and the Aharonov–Casher effect, *i.e.* the dual of the Aharonov–Bohm effect for neutral particles having a magnetic moment (Cimmino *et al.*, 1989).

5.3.7.4. Neutron diffraction topography and other imaging methods

These are the neutron form of the 'topographic' or diffraction imaging techniques, in which an image of a single crystal is obtained through the local variations in Bragg-diffracted intensity due to inhomogeneities in the sample. It is briefly described in Chapter 2.8 of *IT* C. It was pioneered by Doi *et al.* (1971) and by Ando & Hosoya (1972). Like its X-ray counterpart, neutron topography can reveal isolated defects, such as dislocations (Schlenker *et al.*, 1974; Malgrange *et al.*, 1976). Because of the small neutron fluxes available, it is not very convenient for this purpose, since the resolution is poor or the exposure times are very long. On the other hand, the very low absorption of neutrons in most materials makes it quite convenient for observing the gross defect structure in samples that would be too absorbing for X-rays (Tomimitsu & Doi, 1974; Baruchel *et al.*, 1978; Tomimitsu *et al.*, 1983; Kvardakov *et al.*, 1992), or the spatial modulation of distortion due to vibration, for example in quartz (Michalec *et al.*, 1975), and resonant magnetoelastic effects (Kvardakov & Somenkov, 1991). In particular, virtual slices of bulky as-grown samples can be investigated without cutting them using neutron section topography or neutron tomography (Schlenker *et al.*, 1975; Davidson & Case, 1976).

Neutron topography also shows the salient dynamical interference effect, *viz Pendellösung*, visually, in the form of fringes (Kikuta *et al.*, 1971; Malgrange *et al.*, 1976; Tomimitsu & Zeyen, 1978). Its unique feature, however, is the possibility of observing and directly characterizing inhomogeneities in the magnetic structure, *i.e.* magnetic domains of all kinds [ferromagnetic domains (Schlenker & Shull, 1973) and antiferromagnetic domains of various sorts (Schlenker & Baruchel, 1978), including spin-density wave domains (Ando & Hosoya, 1972, 1978; Davidson *et al.*, 1974), 180° or time-reversed domains in some materials and helimagnetic or chirality domains (Baruchel *et al.*, 1990)], or coexisting phases at a first-order phase transition (Baruchel, 1989). In such cases, the contrast is primarily due to local variations in the structure factor, a situation very unusual in X-ray topography, and good crystal quality, leading to dynamical scattering behaviour, is essential in the observation process only in a few cases (Schlenker *et al.*, 1978). It is often crucial, however, for making the domain structure simple enough to be resolved, in particular in the case of antiferromagnetic domains.

Imaging can also be performed for samples that need be neither crystals nor perfect. Phase-contrast imaging of a specimen through which the neutrons are transmitted can be performed in a neutron interferometer. It has been shown to reveal thickness variations by Bauspiess *et al.* (1978) and ferromagnetic domains by Schlenker *et al.* (1980). The same papers showed that phase edges show up as contrast when one of the interferometer paths is blocked, *i.e.* when the sample is placed effectively between perfect, identical crystals set for diffraction in a nondispersive setting. Under the name of neutron radiography with refraction contrast, this technique, essentially a form of Schlieren imaging, was further developed by Podurets, Somenkov & Shil'shtein (1989), Podurets, Somenkov, Chistyakov & Shil'shtein (1989), and Podurets *et al.* (1991), who were able to image internal ferromagnetic domain walls in samples 10 mm thick.

References

Albertini, G., Boeuf, A., Cesini, G., Mazkedian, S., Melone, S. & Rustichelli, F. (1976). *A simple model for dynamical neutron diffraction by deformed crystals. Acta Cryst.* A**32**, 863–868.

Albertini, G., Boeuf, A., Klar, B., Lagomarsino, S., Mazkedian, S., Melone, S., Puliti, P. & Rustichelli, F. (1977). *Dynamical neutron diffraction by curved crystals in the Laue geometry. Phys. Status Solidi A*, **44**, 127–136.

Albertini, G., Boeuf, A., Lagomarsino, S., Mazkedian, S., Melone, S. & Rustichelli, F. (1976). *Neutron properties of curved monochromators. Proceedings of the Conference on Neutron Scattering, Gatlinburg, Tennessee*, USERDA CONF 760601-P2, 1151–1158. Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Al Haddad, M. & Becker, P. J. (1988). *On the statistical dynamical theory of diffraction: application to silicon. Acta Cryst.* A**44**, 262–270.

Ando, M. & Hosoya, S. (1972). *Q-switch and polarization domains in antiferromagnetic chromium observed with neutron-diffraction topography. Phys. Rev. Lett.* **29**, 281–285.

Ando, M. & Hosoya, S. (1978). *Size and behavior of antiferromagnetic domains in Cr directly observed with X-ray and neutron topography. J. Appl. Phys.* **49**, 6045–6051.

Arif, M., Kaiser, H., Clothier, R., Werner, S. A., Berliner, R., Hamilton, W. A., Cimmino, A. & Klein, A. G. (1988). *Fizeau effect for neutrons passing through matter at a nuclear resonance. Physica B*, **151**, 63–67.

Arthur, J. & Horne, M. A. (1985). *Boundary conditions in dynamical neutron diffraction. Phys. Rev. B*, **32**, 5747–5752.

Bacon, G. E. & Lowde, R. D. (1948). *Secondary extinction and neutron crystallography. Acta Cryst.* **1**, 303–314.

Badurek, G., Rauch, H., Wilfing, A., Bonse, U. & Graeff, W. (1979). *A perfect-crystal neutron polarizer as an application of magnetic prism refraction. J. Appl. Cryst.* **12**, 186–191.

Baruchel, J. (1989). *The contribution of neutron and synchrotron radiation topography to the investigation of first-order magnetic phase transitions. Phase Transit.* **14**, 21–29.

Baruchel, J., Guigay, J. P., Mazuré-Espejo, C., Schlenker, M. & Schweizer, J. (1982). *Observation of Pendellösung effect in polarized neutron scattering from a magnetic crystal. J. Phys.* **43**, C7, 101–106.

Baruchel, J., Patterson, C. & Guigay, J. P. (1986). *Neutron diffraction investigation of the nuclear and magnetic extinction in MnP. Acta Cryst.* A**42**, 47–55.

Baruchel, J., Schlenker, M. & Palmer, S. B. (1990). *Neutron diffraction topographic investigations of "exotic" magnetic domains. Nondestruct. Test. Eval.* **5**, 349–367.

Baruchel, J., Schlenker, M., Zarka, A. & Pétroff, J. F. (1978). *Neutron diffraction topographic investigation of growth defects in natural lead carbonate single crystals. J. Cryst. Growth*, **44**, 356–362.

Baryshevskii, V. G. (1976). *Particle spin precession in antiferromagnets. Sov. Phys. Solid State*, **18**, 204–208.

Bauspiess, W., Bonse, U., Graeff, W., Schlenker, M. & Rauch, H. (1978). Result shown in Bonse (1979).

Becker, P. & Al Haddad, M. (1990). *Diffraction by a randomly distorted crystal. I. The case of short-range order. Acta Cryst.* A**46**, 123–129.

Becker, P. & Al Haddad, M. (1992). *Diffraction by a randomly distorted crystal. II. General theory. Acta Cryst.* A**48**, 121–134.

Becker, P. J. & Coppens, P. (1974*a*). *Extinction within the limit of validity of the Darwin transfer equations. I. General formalisms for primary and secondary extinction and their application to spherical crystals. Acta Cryst.* A**30**, 129–147.

Becker, P. J. & Coppens, P. (1974*b*). *Extinction within the limit of validity of the Darwin transfer equations. II. Refinement of extinction in spherical crystals of SrF_2 and LiF. Acta Cryst.* A**30**, 148–153.

Becker, P. J. & Coppens, P. (1975). *Extinction within the limit of validity of the Darwin transfer equations. III. Non-spherical crystals and anisotropy of extinction. Acta Cryst.* A**31**, 417–425.

Belova, N. E., Eichhorn, F., Somenkov, V. A., Utemisov, K. & Shil'shtein, S. Sh. (1983). *Analyse der Neigungsmethode zur Untersuchung von Pendellösungsinterferenzen von Neutronen und Röntgenstrahlen. Phys. Status Solidi A*, **76**, 257–265.

Belyakov, V. A. & Bokun, R. Ch. (1975). *Dynamical theory of neutron diffraction by perfect antiferromagnetic crystals. Sov. Phys. Solid State*, **17**, 1142–1145.

Belyakov, V. A. & Bokun, R. Ch. (1976). *Dynamical theory of neutron diffraction in magnetic crystals. Sov. Phys. Solid State*, **18**, 1399–1402.

Boeuf, A. & Rustichelli, F. (1974). *Some neutron diffraction experiments on curved silicon crystals. Acta Cryst.* A**30**, 798–805.

Bokun, R. Ch. (1979). *Beats in the integrated intensity in neutron diffraction by perfect magnetic crystals. Sov. Phys. Tech. Phys.* **24**, 723–724.

Bonnet, M., Delapalme, A., Becker, P. & Fuess, H. (1976). *Polarised neutron diffraction – a tool for testing extinction models: application to yttrium iron garnet. Acta Cryst.* A**32**, 945–953.

Bonse, U. (1979). *Principles and methods of neutron interferometry*. In *Neutron Interferometry: Proceedings of an International Workshop*, edited by U. Bonse & H. Rauch, pp. 3–33. Oxford: Clarendon Press.

Bonse, U. (1988). *Recent advances in X-ray and neutron interferometry. Physica B*, **151**, 7–21.

Bonse, U. & Hart, M. (1965). *An X-ray interferometer. Appl. Phys. Lett.* **6**, 155–156.

Chakravarthy, R. & Madhav Rao, L. (1980). *A simple method to correct for secondary extinction in polarised-neutron diffractometry. Acta Cryst.* A**36**, 139–142.

Chalupa, B., Michalec, R., Horalik, L. & Mikula, P. (1986). *The study of neutron acoustic effect by neutron diffraction on InSb single crystal. Phys. Status Solidi A*, **97**, 403–409.

Cimmino, A., Opat, G. I., Klein, A. G., Kaiser, H., Werner, S. A., Arif, M. & Clothier, R. (1989). *Observation of the topological Aharonov–Casher phase shift by neutron interferometry. Phys. Rev. Lett.* **63**, 380–383.

Colella, R., Overhauser, A. W. & Werner, S. A. (1975). *Observation of gravitationally induced quantum interference. Phys. Rev. Lett.* **34**, 1472–1474.

Davidson, J. B. & Case, A. L. (1976). *Applications of the fly's eye neutron camera: diffraction tomography and phase transition studies. Proceedings of the Conference on Neutron Scattering, Gatlinburg, Tennessee*, USERDA CONF 760601-P2, 1124–1135. Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Davidson, J. B., Werner, S. A. & Arrott, A. S. (1974). *Neutron microscopy of spin density wave domains in chromium. Proceedings of the 19th Annual Conference on Magnetism and Magnetic Materials*, edited by C. D. Graham and J. J. Rhyne. *AIP Conf. Proc.* **18**, 396–400.

Doi, K., Minakawa, N., Motohashi, H. & Masaki, N. (1971). *A trial of neutron diffraction topography. J. Appl. Cryst.* **4**, 528–530.

Eichhorn, F. (1988). *Perfect crystal neutron optics. Physica B*, **151**, 140–146.

Eichhorn, F., Sippel, D. & Kleinstück, K. (1967). *Influence of oxygen segregations in silicon single crystals on the halfwidth of the double-crystal rocking curve of thermal neutrons. Phys. Status Solidi*, **23**, 237–240.

Guigay, J. P. (1989). *On integrated intensities in Kato's statistical diffraction theory. Acta Cryst.* A**45**, 241–244.

Guigay, J. P. & Chukhovskii, F. N. (1992). *Reformulation of the dynamical theory of coherent wave propagation by randomly distorted crystals. Acta Cryst.* A**48**, 819–826.

Guigay, J. P. & Chukhovskii, F. N. (1995). *Reformulation of the statistical theory of dynamical diffraction in the case E = 0. Acta Cryst.* A**51**, 288–294.

Guigay, J. P. & Schlenker, M. (1979*a*). *Integrated intensities and flipping ratios in neutron diffraction by perfect magnetic crystals*. In *Neutron Interferometry*, edited by U. Bonse & H. Rauch, pp. 135–148. Oxford: Clarendon Press.

Guigay, J. P. & Schlenker, M. (1979*b*). *Spin rotation of the forward diffracted beam in neutron diffraction by perfect magnetic crystals. J. Magn. Magn. Mater.* **14**, 340–343.

Hart, M. & Rodrigues, A. R. D. (1978). *Harmonic-free single-crystal monochromators for neutrons and X-rays. J. Appl. Cryst.* **11**, 248–253.

Hastings, J. B., Siddons, D. P. & Lehmann, M. (1990). *Diffraction broadening and suppression of the inelastic channel in resonant neutron scattering. Phys. Rev. Lett.* **64**, 2030–2033.

Horne, M. A., Finkelstein, K. D., Shull, C. G., Zeilinger, A. & Bernstein, H. J. (1988). *Neutron spin – Pendellösung resonance. Physica B*, **151**, 189–192.

Indenbom, V. L. (1979). *Diffraction focusing of neutrons. JETP Lett.* **29**, 5–8.

International Tables for Crystallography (2004). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by E. Prince. Dordrecht: Kluwer Academic Publishers.

Iolin, E. M. & Entin, I. R. (1983). *Dynamic diffraction of neutrons by high-frequency acoustic waves in perfect crystals. Sov. Phys. JETP*, **58**, 985–989.

Iolin, E. M., Zolotoyabko, E. V., Raïtman, E. A., Kuvdaldin, B. V. & Gavrilov, V. N. (1986). *Interference effects in dynamic neutron diffraction under conditions of ultrasonic excitation. Sov. Phys. JETP*, **64**, 1267–1271.

Kagan, Yu. & Afanas'ev, A. M. (1966). *Suppression of inelastic channels in resonance scattering of neutrons in regular crystals. Sov. Phys. JETP*, **22**, 1032–1040.

Kato, N. (1980*a*). *Statistical dynamical theory of crystal diffraction. I. General formulation. Acta Cryst.* A**36**, 763–769.

Kato, N. (1980*b*). *Statistical dynamical theory of crystal diffraction. II. Intensity distribution and integrated intensity in the Laue cases. Acta Cryst.* A**36**, 770–778.

Kikuta, S., Ishikawa, I., Kohra, K. & Hoshino, S. (1975). *Studies on dynamical diffraction phenomena of neutrons using properties of wave fan. J. Phys. Soc. Jpn*, **39**, 471–478.

Kikuta, S., Kohra, K., Minakawa, N. & Doi, K. (1971). *An observation of neutron Pendellösung fringes in a wedge-shaped silicon single crystal. J. Phys. Soc. Jpn*, **31**, 954–955.

Klar, B. & Rustichelli, F. (1973). *Dynamical neutron diffraction by ideally curved crystals. Nuovo Cimento B*, **13**, 249–271.

Klein, A. G. (1988). *Schrödinger inviolate: neutron optical searches for violations of quantum mechanics. Physica B*, **151**, 44–49.

Klein, A. G. & Werner, S. A. (1983). *Neutron optics. Rep. Prog. Phys.* **46**, 259–335.

Knowles, J. W. (1956). *Anomalous absorption of slow neutrons and X-rays in nearly perfect single crystals. Acta Cryst.* **9**, 61–69.

Korpiun, P. (1966). *Untersuchung ferromagnetischer Strukturen mit einem Zweistrahl-Neutroneninterferometer. Z. Phys.* **195**, 146–170.

Kulda, J. (1988*a*). *The RED extinction model. I. An upgraded formalism. Acta Cryst.* A**44**, 283–285.

Kulda, J. (1988*b*). *The RED extinction model. II. Refinement of extinction and thermal vibration parameters for SrF_2 crystals. Acta Cryst.* A**44**, 286–290.

Kulda, J. (1991). *The RED extinction model. III. The case of pure primary extinction. Acta Cryst.* A**47**, 775–779.

Kulda, J., Baruchel, J., Guigay, J.-P. & Schlenker, M. (1991). *Extinction effects in polarized neutron diffraction from magnetic crystals. I. Highly perfect MnP and YIG samples. Acta Cryst.* A**47**, 770–775.

Kulda, J., Vrána, M. & Mikula, P. (1988). *Neutron diffraction by vibrating crystals. Physica B*, **151**, 122–129.

Kvardakov, V. V., Podurets, K. M., Baruchel, J. & Sandonis, J. (1995). *Precision determination of the structure factors of magnetic neutron scattering from the Pendellösung data. Crystallogr. Rep.* **40**, 330–331.

Kvardakov, V. V., Podurets, K. M., Chistyakov, R. R., Shil'shtein, S. Sh., Elyutin, N. O., Kulidzhanov, F. G., Bradler, J. & Kadečková, S. (1987). *Modification of the domain structure of a silicon–iron single crystal as a result of uniaxial stretching. Sov. Phys. Solid State*, **29**, 228–232.

Kvardakov, V. V. & Somenkov, V. A. (1990). *Observation of dynamic oscillations of intensity of magnetic scattering of neutrons with variation of the orientation of magnetic moments of the sublattices. Sov. Phys. Crystallogr.* **35**, 619–622.

Kvardakov, V. V. & Somenkov, V. A. (1991). *Neutron diffraction study of nonlinear magnetoacoustic effects in perfect crystals of $FeBO_3$ and α-Fe_2O_3. J. Moscow Phys. Soc.* **1**, 33–57.

Kvardakov, V. V. & Somenkov, V. A. (1992). *Magnetic Pendellösung effects in neutron scattering by perfect magnetic crystals. Acta Cryst.* A**48**, 423–430.

Kvardakov, V. V., Somenkov, V. A. & Shil'shtein, S. Sh. (1990*a*). *Observation of dynamic oscillations in the temperature dependence of the intensity of the magnetic scattering of neutrons. Sov. Phys. Solid State*, **32**, 1097–1098.

Kvardakov, V. V., Somenkov, V. A. & Shil'shtein, S. Sh. (1990*b*). *Influence of an orientational magnetic transition in α-Fe_2O_3 on the Pendellösung fringe effect in neutron scattering. Sov. Phys. Solid State*, **32**, 1250–1251.

Kvardakov, V. V., Somenkov, V. A. & Shil'shtein, S. Sh. (1992). *Study of defects in cuprate single crystals by the neutron topography and selective etching methods. Superconductivity*, **5**, 623–629.

Lambert, D. & Malgrange, C. (1982). *X-ray and neutron integrated intensity diffracted by perfect crystals in transmission. Z. Naturforsch. Teil A*, **37**, 474–484.

Malgrange, C., Pétroff, J. F., Sauvage, M., Zarka, A. & Englander, M. (1976). *Individual dislocation images and Pendellösung fringes in neutron topographs. Philos. Mag.* **33**, 743–751.

Mendiratta, S. K. & Blume, M. (1976). *Dynamical theory of thermal neutron scattering. I. Diffraction from magnetic crystals. Phys. Rev.* **14**, 144–154.

Michalec, R., Mikula, P., Sedláková, L., Chalupa, B., Zelenka, J., Petržílka, V. & Hrdlička, Z. (1975). *Effects of thickness-shear vibrations on neutron diffraction by quartz single crystals. J. Appl. Cryst.* **8**, 345–351.

Michalec, R., Mikula, P., Vrána, M., Kulda, J., Chalupa, B. & Sedláková, L. (1988). *Neutron diffraction by perfect crystals excited into mechanical resonance vibrations. Physica B*, **151**, 113–121.

Podurets, K. M., Sokol'skii, D. V., Chistyakov, R. R. & Shil'shtein, S. Sh. (1991). *Reconstruction of the bulk domain structure of silicon iron single crystals from neutron refraction images of internal domain walls. Sov. Phys. Solid State*, **33**, 1668–1672.

Podurets, K. M., Somenkov, V. A., Chistyakov, R. R. & Shil'shtein, S. Sh. (1989). *Visualization of internal domain structure of silicon iron crystals by using neutron radiography with refraction contrast. Physica B*, **156–157**, 694–697.

Podurets, K. M., Somenkov, V. A. & Shil'shtein, S. Sh. (1989). *Neutron radiography with refraction contrast. Physica B*, **156–157**, 691–693.

Rauch, H. (1995). *Towards interferometric Fourier spectroscopy. Physica B*, **213–214**, 830–832.

Rauch, H. & Petrascheck, D. (1978). *Dynamical neutron diffraction and its application*. In *Neutron Diffraction*, edited by H. Dachs, *Topics in Current Physics*, Vol. 6 pp. 305–351. Berlin: Springer.

Rauch, H. & Seidl, E. (1987). *Neutron interferometry as a new tool in condensed matter research. Nucl. Instrum. Methods A*, **255**, 32–37.

Raum, K., Koellner, M., Zeilinger, A., Arif, M. & Gähler, R. (1995). *Effective-mass enhanced deflection of neutrons in noninertial frames. Phys. Rev. Lett.* **74**, 2859–2862.

Scherm, R. & Fåk, B. (1993). *Neutrons*. In *Neutron and Synchrotron Radiation for Condensed Matter Studies (HERCULES Course)*, Vol. 1, edited by J. Baruchel, J. L. Hodeau, M. S. Lehmann, J. R. Regnard & C. Schlenker, pp. 113–143. Les Ulis: Les Editions de Physique and Heidelberg: Springer-Verlag.

Schlenker, M. & Baruchel, J. (1978). *Neutron techniques for the observation of ferro- and antiferromagnetic domains. J. Appl. Phys.* **49**, 1996–2001.

Schlenker, M., Baruchel, J., Perrier de la Bathie, R. & Wilson, S. A. (1975). *Neutron-diffraction section topography: observing crystal slices before cutting them. J. Appl. Phys.* **46**, 2845–2848.

Schlenker, M., Baruchel, J., Pétroff, J. F. & Yelon, W. B. (1974). *Observation of subgrain boundaries and dislocations by neutron diffraction topography. Appl. Phys. Lett.* **25**, 382–384.

Schlenker, M., Bauspiess, W., Graeff, W., Bonse, U. & Rauch, H. (1980). *Imaging of ferromagnetic domains by neutron interferometry. J. Magn. Magn. Mater.* **15–18**, 1507–1509.

Schlenker, M., Linares-Galvez, J. & Baruchel, J. (1978). *A spin-related contrast effect: visibility of 180° ferromagnetic domain walls in unpolarized neutron diffraction topography. Philos. Mag. B*, **37**, 1–11.

Schlenker, M. & Shull, C. G. (1973). *Polarized neutron techniques for the observation of ferromagnetic domains. J. Appl. Phys.* **44**, 4181–4184.

Schmidt, H. H. (1983). *Theoretical investigations of the dynamical neutron diffraction by magnetic single crystals. Acta Cryst.* A**39**, 679–682.

Schmidt, H. H., Deimel, P. & Daniel, H. (1975). *Dynamical diffraction of thermal neutrons by absorbing magnetic crystals. J. Appl. Cryst.* **8**, 128–131.

Sears, V. F. (1978). *Dynamical theory of neutron diffraction. Can. J. Phys.* **56**, 1261–1288.

Shil'shtein, S. Sh., Somenkov, V. A. & Dokashenko, V. P. (1971). *Suppression of (n, γ) reaction in resonant scattering of neutrons by a perfect CdS crystal. JETP Lett.* **13**, 214–217.

Shull, C. G. (1968). *Observation of Pendellösung fringe structure in neutron diffraction. Phys. Rev. Lett.* **21**, 1585–1589.

Shull, C. G. (1986). *Neutron interferometer systems – types and features. Physica B*, **136**, 126–130.

Shull, C. G. & Oberteuffer, J. A. (1972). *Spherical wave neutron propagation and Pendellösung fringe structure in silicon. Phys. Rev. Lett.* **29**, 871–874.

Shull, C. G., Zeilinger, A., Squires, G. L., Horne, M. A., Atwood, D. K. & Arthur, J. (1980). *Anomalous flight time of neutrons through diffracting crystals. Phys. Rev. Lett.* **44**, 1715–1718.

Sippel, D. & Eichhorn, F. (1968). *Anomale inkohärente Streuung thermicher Neutronen bei Bildung stehender Neutronenwellen in nahezu idealen Kristallen von Kaliumdihydrogenphosphat (KDP). Acta Cryst.* A**24**, 237–239.

Sippel, D., Kleinstück, K. & Schulze, G. E. R. (1962). *Nachweis der anomalen Absorption thermischer Neutronen bei Interferenz am Idealkristall. Phys. Status Solidi*, **2**, K104–K105.

Sippel, D., Kleinstück, K. & Schulze, G. E. R. (1964). *Neutron diffraction of ideal crystals using a double crystal spectrometer. Phys. Lett.* **8**, 241–242.

Sippel, D., Kleinstück, K. & Schulze, G. E. R. (1965). *Pendellösungs-Interferenzen mit thermischen Neutronen an Si-Einkristallen. Phys. Lett.* **14**, 174–175.

Sivardière, J. (1975). *Théorie dynamique de la diffraction magnétique des neutrons. Acta Cryst.* A**31**, 340–344.

Somenkov, V. A., Shil'shtein, S. Sh., Belova, N. E. & Utemisov, K. (1978). *Observation of dynamical oscillations for neutron scattering by Ge crystals using the inclination method. Solid State Commun.* **25**, 593–595.

Squires, G. L. (1978). *Introduction to the Theory of Thermal Neutron Scattering*. Cambridge University Press.

Stassis, C. & Oberteuffer, J. A. (1974). *Neutron diffraction by perfect crystals. Phys. Rev. B*, **10**, 5192–5202.

Takagi, S. (1962). *Dynamical theory of diffraction applicable to crystals with any kind of small distortion. Acta Cryst.* **15**, 1311–1312.

Takahashi, T., Tomimitsu, H., Ushigami, Y., Kikuta, S. & Doi, K. (1981). *The very-small angle neutron scattering from neutron-irradiated amorphous silica. Jpn. J. Appl. Phys.* **20**, L837–L839.

Takahashi, T., Tomimitsu, H., Ushigami, Y., Kikuta, S., Doi, K. & Hoshino, S. (1983). *The very-small angle neutron scattering from SiO_2–PbO glasses. Physica B*, **120**, 362–366.

Tasset, F. (1989). *Zero field neutron polarimetry. Physica B*, **156–157**, 627–630.

Taupin, D. (1964). *Théorie dynamique de la diffraction des rayons X par les cristaux déformés. Bull. Soc. Fr. Minéral. Cristallogr.* **87**, 469–511.

Tomimitsu, H. & Doi, K. (1974). *A neutron diffraction topographic observation of strain field in a hot-pressed germanium crystal. J. Appl. Cryst.* **7**, 59–64.

Tomimitsu, H., Doi, K. & Kamada, K. (1983). *Neutron diffraction topographic observation of substructures in Cu-based alloys. Physica B*, **120**, 96–102.

Tomimitsu, H., Takahashi, T., Kikuta, S. & Doi, K. (1986). *Very small angle neutron scattering from amorphous $Fe_{78}B_{12}Si_{10}$. J. Non-Cryst. Solids*, **88**, 388–394.

Tomimitsu, H. & Zeyen, C. (1978). *Neutron diffraction topographic observation of twinned silicon crystal. Jpn. J. Appl. Phys.* **3**, 591–592.

Werner, S. A. (1980). *Gravitational and magnetic field effects on the dynamical diffraction of neutrons. Phys. Rev. B*, **21**, 1774–1789.

Werner, S. A. (1995). *Neutron interferometry tests of quantum theory. Ann. NY Acad. Sci.* **755**, 241–262.

Werner, S. A., Staudenmann, J. L. & Colella, R. (1979). *The effect of the Earth's rotation on the quantum mechanical phase of the neutron. Phys. Rev. Lett.* **42**, 1103–1106.

Zachariasen, W. H. (1967). *A general theory of X-ray diffraction in crystals. Acta Cryst.* **23**, 558–564.

Zeilinger, A. (1995). Private communication.

Zeilinger, A. & Shull, C. G. (1979). *Magnetic field effects on dynamical diffraction of neutrons by perfect crystals. Phys. Rev. B*, **19**, 3957–3962.

Zeilinger, A., Shull, C. G., Horne, M. A. & Finkelstein, K. D. (1986). *Effective mass of neutrons diffracting in crystals. Phys. Rev. Lett.* **57**, 3089–3092.

Zelepukhin, M. V., Kvardakov, V. V., Somenkov, V. A. & Shil'shtein, S. Sh. (1989). *Observation of the Pendellösung fringe effect in magnetic scattering of neutrons. Sov. Phys. JETP*, **68**, 883–886.

Zolotoyabko, E. & Sander, B. (1995). *X-ray diffraction profiles in strained crystals undergoing ultrasonic excitation. The Laue case. Acta Cryst.* A**51**, 163–171.

Author index

Entries refer to chapter number.

Abad-Zapatero, C., 2.3, 3.3
Abdel-Meguid, S. S., 2.3, 3.3
Abi-Ezzi, S. S., 3.3
Ablov, A. V., 2.5
Abrahams, J. P., 2.3
Abrahams, S. C., 2.3, 2.4
Abramowitz, M., 2.1
Achard, M. F., 4.4
Acharya, R., 2.3
Adams, B. W., 5.1
Adams, M. J., 2.3, 2.4
Adams, P. D., 2.3, 2.5
Adamson, R. D., 3.5
Addison, A. W., 2.3
Adiga, P. S., 2.5
Adlhart, W., 4.2
Aeppli, G., 4.2, 4.4
Afanas'ev, A. M., 5.3
Agard, D. A., 2.3, 2.4
Agarwal, R. C., 1.3, 2.4
Agrawal, R. K., 2.5
Aguado, A., 3.5
Aharonov, Y., 3.3
Aharony, A., 4.4
Ahlfors, L. V., 1.3
Ahmed, F. R., 1.3
Åkervall, K., 2.3
Akhiezer, N. I., 1.3
Akimoto, T., 2.3
Akishige, Y., 2.5
Al Haddad, M., 5.3
Alben, R., 4.4
Albertini, G., 5.3
Alden, R. A., 1.3
Alexander, L. E., 4.2
Alexeev, D. G., 4.5
Alford, J. A., 3.5
Al-Khayat, H. A., 4.5
Allegra, G., 2.2
Allen, F. H., 3.3
Allen, L. J., 4.3
Allen, M. P., 3.5
Allinger, N. L., 3.3
Almer, J., 4.2
Als-Nielsen, J., 4.4
Alston, N. A., 1.3
Altermatt, U. D., 1.4
Altmann, S. L., 1.5
Altomare, A., 2.2
Altona, C., 3.3
Alzari, P. M., 5.1
Amador, S., 4.4
Amelinckx, S., 4.3
Amma, E. L., 3.3
Ammon, H. L., 3.5
Amorós, J. L., 4.2
Amorós, M., 4.2
Amos, L. A., 2.5
An, M., 1.3
Anderson, D. C., 3.3
Anderson, D. L., 2.3
Anderson, P. W., 4.4
Anderson, S., 3.3
Andersson, B., 4.3
Andersson, G., 4.4
Ando, M., 5.3
Andreeva, N. S., 2.4
Andrews, J. W., 2.5
Andries, K., 2.3
Angert, I., 2.5
Angress, J. F., 4.1
Anzenhofer, K., 2.2
Apostol, T. M., 1.3
Arai, M., 4.2
Ardito, G., 2.2
Arfken, G., 1.2, 3.4, 3.5
Argos, P., 2.2, 2.3
Arif, M., 5.3
Arley, N., 3.2
Arnautova, Y. A., 3.5
Arndt, U. W., 2.4, 4.2
Arnold, D. B., 3.3
Arnold, E., 2.3
Arnold, H., 1.1, 1.3
Arnott, S., 4.5
Aroyo, M. I., 1.4, 1.5, 4.2
Arrott, A. S., 5.3
Arthur, J., 5.3
Artin, E., 1.3
Ascher, E., 1.3
Ash, J. M., 1.3
Ashcroft, N. W., 1.1
Ashida, T., 2.4
Atkins, E. D. T., 4.5
Atwood, D. K., 5.3
Atwood, J. L., 3.3
Au, A. Y., 2.5, 4.2
Audier, M., 4.2
Auslander, L., 1.3
Authier, A., 1.4, 5.1
Avery, J., 1.2
Avilov, A. S., 2.5
Avraham, D. ben, 4.5
Avrami, M., 2.2
Axe, J. D., 4.2
Axel, F., 4.6
Ayoub, R., 1.3

Bacon, G. E., 5.3
Badasso, M. O., 2.3
Badurek, G., 5.3
Baer, E., 4.5
Bagchi, S. N., 2.3, 4.2, 4.5
Baggio, R., 2.2
Baird, T., 2.5
Bajaj, C., 2.5
Bak, H. J., 2.3
Bak, P., 4.4
Baker, D., 2.3
Baker, E. N., 1.3, 2.3
Baker, T. S., 2.3, 2.5
Baldwin, J. M., 2.5
Baldwin, P. R., 2.5
Balibar, F., 5.1
Ban, N., 2.5
Banaszak, L. J., 2.3
Bancel, P. A., 4.6
Banerjee, K., 2.2
Bannister, C., 2.5
Bansal, M., 1.3
Bantz, D., 1.3
Bar, J., 2.5
Barakat, R., 1.3, 2.1
Barbour, L. J., 3.3
Bardhan, P., 4.2
Barham, P. J., 4.5
Barnea, Z., 1.2
Barnes, W. H., 1.3
Barois, P., 4.4
Barrett, A. N., 1.3
Barrington Leigh, J., 4.5
Barry, C. D., 3.3
Bartels, K., 1.3, 2.3, 2.4
Baruchel, J., 5.3
Baryshevskii, V. G., 5.3
Barzaghi, M., 3.5
Basett-Jones, D. P., 2.5
Bash, P. A., 3.3
Baskaran, S., 4.5
Basokur, A. T., 2.5
Batterman, B. W., 5.1
Baturic-Rubcic, J., 4.4
Bauer, G., 4.2
Bauer, J., 3.5
Bauer, P., 3.5
Baumeister, W., 2.5
Bauspiess, W., 5.3
Baxter, W., 2.5
Bazterra, V. E., 3.5
Beaglehole, D., 4.4
Bean, A. J., 2.5
Beck, T. L., 3.5
Becker, P. J., 1.2, 5.3
Beckmann, E., 2.5
Beddell, C., 3.3
Bedzyk, M. J., 5.1
Beer, T. de, 2.5
Beevers, C. A., 1.3, 2.3
Beintema, J. J., 2.3
Bellamy, H. D., 3.3
Bellamy, K., 4.5
Bellard, S., 3.3
Bellen, H., 2.5
Bellissent, R., 4.2
Bellman, R., 1.3
Bellocq, A. M., 4.4
Bellon, P. L., 2.5
Belova, N. E., 5.3
Belyakov, V. A., 5.3
Benattar, J. J., 4.4
Bender, R., 2.5
Bendersky, L. A., 2.5
Bengtsson, U., 2.3
Benguigui, L., 4.4
Bennett, D. W., 3.3
Bennett, J. M., 1.3
Bensimon, D., 4.4
Bentley, J., 1.2
Berberian, S. K., 1.3
Berendzen, J., 2.3
Berger, J. E., 2.5
Bergman, G., 3.2
Bergmann, K., 4.4
Berkowitz, M. L., 3.5
Berliner, R., 5.3
Berman, H. M., 1.4
Bern, M., 2.5
Bernstein, F. C., 3.3
Bernstein, H. J., 5.3
Bernstein, J., 3.5
Bernstein, S., 2.1
Berry, M. V., 5.2
Bertaut, E. F., 1.3, 1.4, 2.2, 3.4
Berthold, T., 4.2
Berthou, J., 2.3
Bessière, M., 4.2
Bethe, H. A., 1.2, 2.5, 5.2
Bethge, P. H., 3.3
Beurskens, G., 2.2
Beurskens, P. T., 2.2, 2.3
Beyeler, H. U., 4.2
Bhat, T. N., 2.3, 2.4
Bhattacharya, R. N., 1.3
Bhuiya, A. K., 2.2
Bieberbach, L., 1.3
Biemond, J., 2.5
Bienenstock, A., 1.3, 1.4
Bijvoet, J. M., 2.2, 2.3, 2.4
Bilderback, D. H., 5.1
Billard, J., 4.4
Billard, L., 4.6
Billinge, S. J. L., 4.2
Bilz, H., 4.1
Bing, D. H., 3.3
Bird, D., 2.5
Birgeneau, R. J., 4.4
Biswas, A., 4.5
Blackman, M., 2.5, 4.1, 5.2
Blackwell, J., 4.5
Blahut, R. E., 1.3
Blake, A. J., 3.3
Blanc, E., 2.3
Blaser, H., 2.3
Blech, I., 2.5, 4.6
Bleistein, N., 1.3
Blessing, R. H., 2.2
Bley, F., 4.2
Blinc, R., 4.4
Bloch, F., 1.1, 4.1
Bloomer, A. C., 1.3, 2.3, 3.3
Blow, D. M., 1.3, 2.2, 2.3, 2.4
Bluhm, M. M., 2.3
Blume, J., 5.2
Blume, M., 5.3
Blundell, D. J., 4.5
Blundell, T. L., 2.3, 2.4
Bochner, S., 1.3
Bode, W., 1.3, 2.4
Bodo, G., 2.3
Boer, J. L. de, 4.2
Boerrigter, S. X. M., 3.5
Boettger, J. C., 3.5
Boeuf, A., 5.3
Böhm, H., 4.6
Bohm, J., 2.5
Böhme, R., 2.2
Böhringer, M., 5.1
Boisset, N., 2.5
Boissieu, M. de, 4.2
Bokhoven, C., 2.4
Bokun, R. Ch., 5.3
Bommel, A. J. van, 2.3, 2.4
Bondot, P., 1.3
Böni, P., 4.2
Bonneau, P. R., 2.3
Bonnet, M., 5.3
Bono, P. R., 3.3
Bonse, U., 5.1, 5.3
Boon, M., 1.5
Booth, A. D., 1.3
Boots, B., 2.5
Borell, A., 2.3
Borie, B., 4.2
Born, M., 1.2, 1.3, 4.1, 5.1, 5.2
Borrmann, G., 5.1
Bosman, W. P., 2.2
Botha, J. D., 3.3
Böttcher, B., 2.5
Böttcher, C. J. F., 3.5
Boublik, M., 2.5
Bouckaert, L. P., 1.5
Boudard, M., 4.2
Boulay, D. J. du, 1.4, 2.2
Bouman, J., 2.2
Bourne, P. E., 1.4
Boutin, H., 4.1
Bouwman, W. G., 4.4
Bowen, D. K., 5.1
Bown, M., 4.2
Boyd, D. B., 3.3
Boyer, L., 4.1
Boyer, P. L., 2.3
Boyle, L. L., 1.5
Boysen, H., 4.2
Bracewell, R. N., 1.3, 2.5
Bradler, J., 5.3
Bradley, A. J., 2.4
Bradley, C. J., 1.5, 4.2
Bragg, L., 1.4
Bragg, W. H., 1.3
Bragg, W. L., 1.3, 2.3, 5.1
Braig, K., 2.5
Brämer, R., 4.2
Brand, P., 4.4
Brandenburg, N. P., 3.3
Braslau, A., 4.4
Braun, H., 3.3
Braun, P. B., 2.3
Bremermann, H., 1.3
Bremmer, H., 1.3
Brennan, S., 5.1
Brice, M. D., 1.3, 3.3
Bricogne, G., 1.3, 2.2, 2.3, 2.5, 3.3, 4.5
Brigham, E. O., 1.3
Brill, R., 1.3
Brinkman, W. F., 4.4

AUTHOR INDEX

Brisbin, D., 4.4
Brisse, F., 2.5, 4.5
Britten, P. L., 1.3, 2.2
Broach, R. W., 4.2
Brock, J. D., 4.4
Brockhouse, B. N., 4.1
Broderson, S., 3.5
Brooks, B. R., 3.3
Brooks, C. L. III, 2.5
Brooks, J. D., 4.4
Brown, C. J., 4.5
Brown, D., 3.5
Brown, F., 2.3
Brown, G. S., 4.4
Brown, H., 1.3
Brown, I. D., 1.4
Brown, M. D., 3.3
Brucolleri, R. E., 3.3
Bruijn, N. G. de, 1.3
Bruins, E. M., 4.5
Bruins Slot, H. J., 2.2
Bruinsma, R., 4.4
Brümmer, O., 5.1
Brunger, A. T., 2.2, 2.3
Brünger, A. T., 1.3, 2.3, 4.5
Bruno, I. J., 3.3
Bryan, R. K., 1.3, 4.5
Bu, X., 4.2, 4.6
Bubeck, E., 4.2
Buch, K. R., 3.2
Budai, J., 4.4
Budinger, T. F., 2.5
Buehner, M., 2.3
Buerger, M. J., 1.1, 1.4, 2.2, 2.3
Bujosa, A., 1.3
Bullough, R. K., 2.3
Bülow, R., 1.3
Bunkóczi, G., 2.3
Bunn, C. W., 4.5
Bunshaft, A. J., 3.3
Burandt, B., 4.2, 4.6
Burch, S. F., 1.3
Burd, C. G., 2.5
Burdina, V. I., 2.3
Burgess, W. G., 2.5
Bürgi, H.-B., 3.3, 4.2
Burkel, E., 4.1
Burkert, U., 3.3
Burkov, S., 4.6
Burla, M. C., 2.2
Burnett, M. N., 3.3
Burnett, R. M., 1.3, 2.3
Burnside, W., 1.3
Burrus, C. S., 1.3
Busetta, B., 2.2
Busing, W. R., 1.3, 3.1, 3.4
Bussler, P., 2.5
Butcher, S. J., 2.5
Butler, B. D., 4.2
Buttle, K., 2.5
Buxton, B., 2.5, 5.2
Buyers, W. J. L., 4.1
Byerly, W. E., 1.3
Byler, M. A., 4.5
Bystroff, C., 2.3

Cael, J. J., 4.5
Caglioti, G., 4.2
Cahn, J. W., 2.5, 4.6
Caillé, A., 4.4
Calabrese, G., 2.2
Caliandro, R., 2.2
Calvayrac, Y., 4.2
Camalli, M., 2.2
Cambillau, C., 3.3
Campagnari, F., 4.5
Campbell, B. J., 4.2
Campbell, G. A., 1.3
Campbell Smith, P. J., 4.5
Cannillo, E., 2.4
Canright, G. S., 4.2
Cantino, M., 4.5
Capillas, C., 1.4, 1.5
Carathéodory, C., 1.3
Carazo, J. M., 2.5
Cardona, M., 4.1
Carlile, C. J., 4.4
Carlisle, C. H., 2.3
Carlson, J. M., 4.4
Carragher, B., 2.5
Carrascosa, J. L., 2.5
Carroll, C. E., 4.5
Carrozzini, B., 2.2
Carslaw, H. S., 1.3
Cartan, H., 1.3
Carter, R. E., 3.3
Cartwright, B. A., 3.3
Cascarano, G. L., 2.2
Case, A. L., 5.3
Casher, A., 1.5
Caspar, D. L. D., 2.3, 4.5
Castellano, E. E., 2.2
Cavicchi, E., 3.3
Cenedese, P., 4.2
Cesini, G., 5.3
Ceska, T. A., 2.5
Chacko, K. K., 2.4
Chakravarthy, R., 5.3
Challacombe, M., 3.5
Challifour, J. L., 1.3
Chalupa, B., 5.3
Champeney, D. C., 1.3
Champness, J. N., 1.3, 2.3, 3.3
Champness, N. R., 3.3
Chan, A. S., 2.5
Chan, D. S. K., 1.3
Chan, K. K., 4.4
Chandrasekaran, R., 4.5
Chandrasekhar, S., 4.4
Chang, G., 2.3
Chang, S.-L., 5.1
Chanzy, H., 2.5, 4.5
Chao, Y., 2.5
Chao-de, Z., 2.2
Chaplot, S. L., 4.1
Chapman, D., 4.4
Chapman, M. S., 2.3
Chapon, L. C., 3.3
Chapuis, G., 4.6
Charvolin, J., 4.4
Chataka, T., 4.2
Cheetham, A. K., 4.2
Chemburkar, S. R., 3.5
Chen, H. S., 4.2
Chen, J., 2.5
Chen, J. H., 4.4
Chen, S., 2.5
Cheng, N., 2.5
Cheng, R. H., 2.5
Cheng, T. Z., 2.5
Cherns, D., 2.5
Chew, M., 4.5
Chiang, L. Y., 4.4
Chin, C., 2.5
Chisholm, J., 3.5
Chistyakov, R. R., 5.3
Chiu, S. N., 2.5
Chiu, W., 2.5
Chivers, R. A., 4.5
Choi, H.-K., 2.3
Choplin, F., 3.3
Chow, M., 2.3
Christensen, F., 4.4
Christian, P. E., 2.5
Chu, K. C., 4.4
Chukhovskii, F. N., 5.1, 5.3
Church, G. M., 1.3, 2.4
Churchill, R. V., 1.3
Cimmino, A., 5.3
Cisarova, I., 4.2, 4.6
Cisneros, G. A., 3.5
Civalleri, B., 3.5
Clapp, P. C., 4.3
Clark, A. D. Jr, 2.3
Clark, E. S., 4.5
Clark, J. J., 2.5
Clark, N. A., 4.4
Clark, P., 2.3
Clarke, P. J., 4.2
Clarke, R., 4.4
Clastre, J., 2.3
Clausen, K. N., 4.2
Clementi, E., 1.2
Clews, C. J. B., 1.3, 3.2
Clore, G. M., 2.3
Clothier, R., 5.3
Coates, D., 4.4
Cochran, W., 1.1, 1.2, 1.3, 1.4, 2.2, 2.3, 2.5, 3.2, 4.1, 4.5
Cockcroft, J. K., 3.3
Cockrell, P. R., 3.3
Cohen, D., 2.5
Cohen, J. B., 4.2
Cohen, N. C., 3.3
Cohen-Tannoudji, C., 1.2
Cole, H., 5.1
Cole, J. C., 3.3
Colella, R., 5.3
Colin, P., 3.3
Collar, A. R., 5.2
Coller, E., 2.2, 2.3
Collett, J., 4.4
Collins, D. M., 1.3, 2.2, 2.3, 2.4, 3.3
Collongues, R., 4.2
Colman, P. M., 1.3, 2.3
Comarmond, M. B., 2.3
Comes, R., 4.2
Condon, E. V., 1.2
Connell, S. R., 2.5
Connolly, M. L., 3.3
Conradi, E., 4.2
Convert, P., 4.2
Conway, J. F., 2.5
Cooley, J. W., 1.3
Cooper, M. J., 4.2
Coppens, P., 1.2, 3.5, 4.2, 4.6, 5.3
Cordes, A. W., 3.2
Cordingley, M. G., 2.3
Corey, R. B., 1.3, 2.3
Corfield, P. W. R., 2.3
Cork, J. M., 2.4
Corones, J., 5.2
Coster, D., 2.4
Cotton, F. A., 3.3
Couch, G. S., 2.5
Coulson, C. A., 1.2
Coulter, C. L., 2.2
Courville, D. A., 2.3
Courville-Brenasin, J. de, 4.2
Cowan, P. L., 5.1
Cowan, S. W., 2.5, 3.3
Cowley, J. M., 2.5, 4.2, 4.3, 4.5, 5.2
Cowtan, K., 2.3
Cox, E. G., 1.3
Cox, J. M., 2.3, 2.4
Cox, S. R., 3.5
Coxeter, H. S. M., 1.3
Cracknell, A. P., 1.5, 4.2
Cramér, H., 1.3, 2.1, 2.5
Cramer, R. III, 3.3
Craven, B. M., 3.5
Crick, F. H. C., 1.3, 2.2, 2.3, 2.4, 2.5, 4.5
Cromer, D. T., 2.3, 2.4
Crooker, P. P., 4.4
Crowfoot, D., 2.3
Crowther, R. A., 1.3, 2.2, 2.3, 2.5, 4.5
Crozier, P. A., 2.5
Cruickshank, D. W. J., 1.2, 1.3, 2.4
Crutchfield, J. P., 4.2
Cullen, D. L., 3.3
Cullis, A. F., 2.3, 2.4
Culver, J. N., 4.5
Cummins, H. Z., 4.6
Cummins, P. G., 3.4
Cunningham, D., 3.3
Currat, R., 4.2
Curtis, C. W., 1.3
Curtis, R. J., 4.4
Cutfield, J. F., 2.2
Czaplewski, C., 3.5
Czaplewsky, C., 3.5
Czerwinski, E. W., 2.3
Dabrowski, M., 2.5
Dai, J.-B., 2.3
Dale, D., 2.4
Dam, A. van, 3.3
Dana, S. S., 4.4
Daniel, H., 5.3
Daniels, H. E., 1.3
Darby, G., 2.3
Darden, T. A., 3.5
Dark, R., 3.3
Darwin, C. G., 5.1
Das, K., 2.3
Dasgupta, C., 4.4
D'Astuto, M., 4.1
Daubeny, R. de P., 4.5
Dauter, M., 2.3
Dauter, Z., 2.3
Davey, S. C., 4.4
Davidov, D., 4.4
Davidson, E. R., 3.5
Davidson, J. B., 5.3
Davidson, W., 2.3
Davies, B. L., 1.5
Davies, D. R., 1.3
Davis, M. E., 2.3
Davis, P. J., 3.4
Dawson, B., 1.2, 2.5
Day, D., 4.5
Day, G. M., 3.5
Dayringer, H. E., 3.3
De Caro, L., 2.2
De Facio, B., 5.2
De Gennes, P. G., 4.4
De Hoff, R., 4.4
De Meulenaare, P., 4.3
De Ridder, R., 4.3
De Titta, G. T., 2.2, 2.5, 4.5
De Vries, H. L., 4.4
Dea, I. C. M., 4.5
Debaerdemaeker, T., 2.2
DeBoissieu, M., 4.6
Debreczeni, J. É., 2.3
Debye, P., 4.1
Declercq, J.-P., 2.2
Dederichs, P. H., 4.2
Deem, M. W., 3.5
Dehlinger, U., 4.6
Deimel, P., 5.3
Deisenhofer, J., 1.3, 2.3, 2.4
DeLano, W. L., 2.3
Delapalme, A., 5.3
Delaunay, B., 1.5
DeLeeuw, S. W., 3.5
Della Valle, R. G., 3.5
Dellby, N., 4.3
Deming, K., 3.5
Deming, W. E., 3.2
Dempsey, S., 3.3
Demus, D., 4.4
Denny, R., 4.5
Denson, A. K., 3.3
DeRosier, D. J., 2.5, 4.5
Deserno, M., 3.5
Destrade, C., 4.4
Destro, R., 3.5
DeTitta, G. T., 2.2, 2.5, 4.5
Deutsch, M., 4.4
Dewar, R. B. K., 2.2
DeWette, F. W., 3.4
Diamond, R., 1.3, 3.3, 4.5
Dickerson, R. E., 1.3, 2.2, 2.3, 2.4
Diele, S., 4.4
Dietrich, H., 1.3
Dieudonné, J., 1.3
Dijkstra, B. W., 3.3
Dilanian, R. A., 3.3
Dimon, P., 4.4
Ding, D., 4.2
Ding, J., 2.3
Dintzis, H. M., 2.3
Dirac, P. A. M., 1.3
Dirl, R., 1.5
DiSalvo, F. J., 4.4
Ditchfield, R., 1.2

AUTHOR INDEX

Diu, B., 1.2
Djurek, D., 4.4
Dobrott, R. D., 2.3
Dodson, E., 2.2, 2.3, 2.4
Dodson, E. J., 1.3, 2.2, 2.3, 3.3
Dodson, G. G., 2.2, 2.3, 3.3
Doerschuk, P. C., 2.5
Doesburg, H. M., 2.2
Doi, K., 5.3
Dokashenko, V. P., 5.3
Dolata, D. P., 3.3
Dolling, C., 4.1
Dolling, G., 4.2
Dolomanov, O. V., 3.3
Domany, E., 4.4
Donabauer, J., 3.5
Dong, W., 2.5
Donohue, J., 1.3, 2.3
Donovan, B., 4.1
Dorna, V., 4.2
Dorner, B., 4.2
Dorner, C., 4.2
Dorset, D. L., 2.5, 4.5
Doubleday, A., 3.3
Doucet, J., 4.4
Douglas, A. S., 2.2
Dovesi, R., 3.5
Dowell, W. C. T., 2.5
Downing, K. H., 2.5
Downs, R. T., 3.3
Dowty, E., 1.4, 3.3
Doyle, P. A., 4.3, 4.5
Dräger, J., 4.6
Drenth, J., 4.5
Drits, V. A., 2.5
Dror, R. O., 3.5
Duan, X., 2.5
Duane, W., 1.3
Dube, P., 2.5
Dubernat, J., 4.2
Dubois, J. C., 4.4
Duce, D. A., 3.3
Duijneveldt, F. B. van, 3.5
Dumrongrattana, S., 4.4
Duncan, W. J., 5.2
Dunitz, J. D., 1.2, 3.5
Dunmur, D. A., 3.4
Durrant, J. L. A., 4.4
Dvoryankin, V. F., 2.5
D'yakon, I. A., 2.5
Dym, H., 1.3
Dyott, T. M., 3.3
Dziki, W., 3.5
Dzyabchenko, A., 3.5

Eades, J. A., 2.5
Eaglesham, D. J., 2.5
Eaker, D., 2.3
Eastwood, J. W., 3.5
Eastwood, M. P., 3.5
Eckold, G., 4.2
Edgington, P. R., 3.3
Edmonds, J. W., 2.2, 2.5, 4.5
Edwards, O. S., 4.2
Egami, T., 4.2
Egelman, E., 2.5
Egert, E., 2.2, 2.3
Eichhorn, F., 5.3
Eijck, B. P. van, 3.5
Eiland, P. F., 1.3
Einstein, A., 4.1
Einstein, J. E., 2.4
Eisenberg, D., 2.2, 2.3, 2.4
Eklundh, J. O., 1.3
Elder, M., 4.5
Eliopoulos, E. E., 3.3
Eller, G. von, 2.2
Elyutin, N. O., 5.3
Emery, V. J., 4.2
Emr, S. D., 2.5
Enderle, G., 3.3
Endoh, H., 2.5
Endoh, Y., 2.5
Endres, H., 4.2
Engel, G., 3.5
Engel, P., 1.3
Englander, M., 5.3
Entin, I. R., 5.3
Epstein, J., 4.2
Erdélyi, A., 1.3
Erickson, H. P., 2.5
Erickson, J. W., 2.3
Erk, P., 3.5
Eschenbacher, P. W., 1.3
Esnouf, R., 2.3
Essmann, U., 3.5
Estermann, M., 4.2, 4.6
Etheridge, J., 5.2
Etherington, G., 4.4
Evans, A. C., 3.3
Evans, G., 2.3
Evans, N. S., 2.5
Evans, P., 2.4
Evans, P. R., 3.3
Evans-Lutterodt, K. W., 4.4
Evjen, H. M., 3.4
Ewald, P. P., 1.1, 1.3, 1.4, 3.4, 5.1
Exelby, D. R., 2.5
Eyges, L., 3.5

Faber, T. E., 4.4
Facelli, J. C., 3.5
Faetti, S., 4.4
Faggiani, R., 2.1
Fåk, B., 5.3
Fan, C. P., 4.4
Fan, H.-F., 2.2, 2.5
Farach, H. A., 3.3
Farber, A. S., 4.4
Farkas, D. R., 1.3
Farrants, G., 1.4, 2.5
Farrants, G. W., 3.3
Farrow, N. A., 2.5
Farrugia, L. J., 3.3
Favin, D. L., 1.3
Fayard, M., 4.2
Fedotov, A. F., 2.5
Fehlhammer, H., 2.3
Feig, E., 1.3
Feil, D., 1.2
Feiner, S. K., 3.3
Feldkamp, L. A., 4.1
Feldmann, R. J., 3.3
Fellmann, D., 2.5
Feltynowski, A., 2.5
Fender, B. E. F., 4.2
Fernandez, J.-J., 2.5
Ferrara, J. D., 2.3
Ferraris, G., 2.5
Ferrarro, M. B., 3.5
Ferrel, R. A., 4.3
Ferrin, T. E., 2.5, 3.3
Fewster, P. F., 5.1
Feynman, R., 5.2
Fields, P. M., 4.3
Fillipini, G., 3.5
Filman, D. J., 2.3
Finch, J. T., 2.5
Finger, L. W., 3.3
Fingerland, A., 5.1
Finkelstein, K. D., 5.3
Finkenstadt, V. L., 4.5
Finn, R., 2.5
Fischer, J., 2.3
Fischer, K., 1.2
Fischer, R. X., 3.3
Fischer, W., 1.4
Fisher, J., 2.2
Fisher, P. M. J., 4.3
Fishman, G., 2.5
Fiske, S. J., 2.2
Fitzgerald, P. M. D., 1.4, 2.3
Flack, H. D., 4.2
Flautt, T. J., 4.4
Fleming, R. M., 4.4
Flensburg, C., 2.3
Fletterick, R. J., 2.3, 3.3
Flook, R. J., 2.4
Foadi, J., 2.2
Fock, R., 5.2
Fogel, D. B., 2.3
Foley, J. D., 3.3
Folkhard, W., 1.3
Fontaine, D. de, 4.2
Fontell, K., 4.4
Fontes, E., 4.4, 5.1
Ford, G. C., 2.3
Ford, L. O., 3.3
Fornberg, A., 1.3
Forst, R., 4.2
Förster, E., 5.1
Forster, F., 2.5
Forsyth, J. B., 1.3, 4.2
Forsyth, V. T., 4.5
Fortier, S., 2.2
Fortuin, C. M., 3.4
Forwood, C. T., 2.5
Foster, R. M., 1.3
Foucher, P., 4.4
Fouret, P., 4.2
Fouret, R., 4.2
Fourme, R., 2.3
Fout, G. S., 2.3
Fowler, R. H., 1.3
Fowweather, F., 1.3
Fox, G., 2.3
Frampton, C. S., 3.5
Francis, N., 2.5
Frangakis, A. S., 2.5
Frank, F. C., 4.4
Frank, J., 2.5
Frankenberger, E. A., 2.3
Franklin, R. E., 4.5
Franulovic, K., 4.4
Franx, M., 2.1
Fraser, D., 2.5
Fraser, R. D. B., 4.5
Frazer, R. A., 5.2
Freeman, A. J., 1.2, 4.3
Freeman, H. C., 2.4
Freer, A. A., 2.2
Freer, S. T., 1.3, 2.4
Freiser, M. J., 4.4
French, A. D., 4.5
French, S., 2.1, 2.2, 2.4
Frenkel, D., 3.5
Frey, F., 4.2
Frey, S., 2.5
Fridborg, K., 2.3
Fridrichsons, J., 2.3
Friedel, G., 1.3, 4.4
Friedlander, F. G., 1.3
Friedlander, P. H., 1.3
Friedman, A., 1.3
Frobenius, G., 1.3
Frost, J. C., 4.4
Frost-Jensen, A., 4.2, 4.6
Fry, E., 2.3
Fryer, J. R., 2.5, 4.5
Fuess, H., 5.3
Fujii, Y., 4.1
Fujimoto, F., 2.5, 4.3, 5.2
Fujinaga, M., 2.3
Fujiwara, A., 2.5
Fujiwara, K., 2.5, 5.2
Fujiyoshi, Y., 2.5
Fukuhara, A., 2.5, 5.2
Fukuyama, K., 2.3
Fuller, S. D., 2.5
Fuller, W., 4.5
Fultz, B., 4.2
Fung, K. K., 2.5
Furey, W., 2.3
Furie, B., 3.3
Furie, B. C., 3.3
Furusaka, M., 4.2
Fusti-Molnar, L., 3.5

Gabashvili, I. S., 2.5
Gabor, D., 2.5
Gähler, R., 5.3
Galerne, Y., 4.4
Gallo, L., 3.3
Gallo, S. M., 2.2
Gallop, J. R., 3.3
Gallwitz, U., 4.5
Gane, P. A. C., 4.4
Gannon, M. G. J., 4.4
Garcia, A. E., 3.5
Garcia-Golding, F., 4.4
Garcia-Granda, S., 2.2
García-Rodríguez, L., 2.2
Gardner, K. H., 4.5
Garland, C. W., 4.4
Garland, Z. G., 4.4
Garman, E., 2.3
Garrido, J., 2.3
Gasparoux, H., 4.4
Gassmann, J., 1.3, 2.3, 2.4, 2.5
Gatti, M., 4.4
Gaughan, J. P., 4.4
Gautier, F., 4.2
Gavezzotti, A., 3.5
Gavrilov, V. N., 5.3
Gay, R., 2.3
Gaykema, W. P. J., 2.3
Gayle, F. W., 4.2
Gebhard, W., 4.5
Geddes, A. J., 3.3
Gehlen, P., 4.2
Gehlhaar, D. K., 2.3
Gehring, K., 2.5
Geil, P. H., 4.5
Geisel, T., 4.2
Gelder, R. de, 2.2
Gel'fand, I. M., 1.3
Geller, M., 3.3
Gentleman, W. M., 1.3
Georgopoulos, P., 4.2
Gerhard, O. E., 2.1
Gerlach, P., 4.2
Germain, G., 2.2, 2.5
Germian, C., 4.4
Gerold, V., 4.2
Giacovazzo, C., 2.1, 2.2
Giarrusso, F. F., 3.3
Gibbons, P. C., 4.2
Gibbs, J. W., 2.3
Gibson, M. A., 2.5
Giegé, R., 2.3
Giesebrecht, J., 2.5
Gilbert, P. F. C., 2.5
Gill, P. E., 3.3
Gill, P. M. W., 3.5
Gillan, B. E., 4.2
Gilli, G., 2.4
Gilligan, K., 3.3
Gilliland, G. L., 3.3
Gillis, J., 1.3, 2.2
Gilmore, C. J., 2.2, 2.5, 4.5
Gilmore, R., 5.2
Gingrich, N. S., 1.3
Girling, R. L., 3.3
Gjønnes, J., 2.5, 4.3, 5.2
Gjønnes, K., 2.5
Glaeser, R. M., 2.5
Glasser, M. L., 3.4
Glatigny, A., 3.3
Glauber, R., 2.5
Glazer, A. M., 4.2
Glosli, J., 3.5
Glück, M., 1.5
Glucksman, M. J., 4.5
Glykos, N. M., 2.3
Go, N., 3.3
Goddard, T. D., 2.5
Godréche, C., 4.6
Goedkoop, J. A., 1.3, 2.2
Goff, J. P., 4.2
Golas, M. M., 2.5
Goldman, A. I., 4.2, 4.6
Goldstine, H. H., 1.3
Golovchenko, J. A., 5.1
Golub, G. H., 2.5
Goncharov, A. B., 2.5
Gonzalez, A., 4.5

AUTHOR INDEX

Gonzalez, R. F., 2.5
Good, I. J., 1.3
Goodby, J. W., 4.4
Goodman, P., 2.5, 5.2
Goodyear, G., 2.5
Goossens, D. J., 4.2
Gordon, R., 2.5
Gosling, R. G., 4.5
Gossling, T. H., 3.3
Gould, R. O., 2.2
Gouyet, J. F., 4.6
Gowen, B., 2.5
Graaf, H. de, 4.5
Graaff, R. A. G. de, 2.2
Graeff, W., 5.1, 5.3
Gragg, J. E., 4.2
Gramlich, V., 2.2
Gransbergen, E. F., 4.4
Grant, D. F., 2.2
Grassucci, R. A., 2.5
Gratias, D., 2.5, 4.6, 5.2
Grau, U. M., 2.3
Gray, G. W., 4.4
Green, D. W., 2.3, 2.4
Green, E. A., 2.2, 2.4
Greenall, R. J., 4.5
Greenberg, W. L., 2.5
Greenblatt, D. M., 2.5
Greenhalgh, D. M. S., 1.3
Greer, J., 3.3
Grems, M. D., 1.3
Grenander, U., 1.3
Griffith, J. P., 2.3
Grigorieff, N., 2.5
Grimm, H., 1.3, 4.2
Grindley, J., 3.5
Grinstein, G., 4.4
Gronsky, R., 2.5
Gros, P., 2.3
Gross, L., 1.3
Grosse-Kunstleve, R. W., 1.4, 2.2, 2.3
Grubb, D. T., 4.5
Grzegory, I., 4.1
Guagliardi, A., 2.2
Gubbens, A. J., 4.3
Guessoum, A., 1.3
Guigay, J. P., 5.3
Guillon, D., 4.4
Guinier, A., 4.2
Gull, S. F., 1.3
Gullberg, G. T., 2.5
Gunther, L., 4.4
Gur, Y., 1.5
Gurskaya, G. V., 2.5
Guru Row, T. N., 1.2
Guryan, C. A., 4.2
Gutierrez, G. A., 4.5
Guyot-Sionnest, P., 4.4

Haas, F. de, 2.5
Hadamard, J., 1.3
Haeffner, D. R., 4.2
Haefner, K., 4.2
Haibach, T., 4.6
Hall, I. H., 4.5
Hall, M., 1.3
Hall, R. J., 2.5
Hall, S. R., 1.4, 2.2
Halla, F., 4.2
Hall-Wallace, M., 3.3
Halperin, B. I., 4.4
Hamaker, C., 2.5
Hamilton, W. A., 5.3
Hamilton, W. C., 2.3, 2.4, 3.1, 3.2, 4.5
Hancock, H., 2.2
Handelsman, R. A., 1.3
Hansen, J., 2.5
Hansen, N. K., 1.2
Hao, Q., 2.2
Harada, J., 4.2
Harada, Y., 2.3, 2.5, 4.3
Harauz, G., 2.5
Harburn, G., 4.2
Harding, M. M., 2.2, 2.3
Hardman, K. D., 3.3
Hardouin, F., 4.4
Hardy, G. H., 1.3
Harford, J., 4.5
Harker, D., 1.3, 2.1, 2.2, 2.3, 2.4
Harrington, M., 2.3
Harris, D. B., 1.3
Harris, G. W., 4.2
Harris, M. R., 3.3
Harrison, S. C., 1.3, 2.3
Harrison, W. A., 4.1
Hart, M., 5.1, 5.3
Hart, R. G., 2.4
Hartman, P., 1.3
Hartree, D. R., 1.2
Hartsuck, J. A., 2.3
Hasegawa, K., 4.5
Haseltine, J. H., 4.4
Hashimoto, H., 2.5
Hashimoto, S., 4.2, 4.3
Hashimoto, T., 2.5
Hass, B. S., 3.3
Hastings, C. Jr, 3.4
Hastings, J. B., 5.3
Hata, Y., 2.4
Hatch, D. M., 1.5
Haubold, H. G., 4.2
Hauptman, H., 1.3, 2.1, 2.2, 2.3, 2.4, 2.5, 4.5
Hausdorff, F., 4.6
Hauser, J., 3.3
Havelka, W., 2.5
Havighurst, R. J., 1.3
Hayakawa, M., 4.2
Hayes, W., 4.2
Hazen, E. E., 3.3
He, L. X., 2.5
He, Y., 2.3
Head-Gordon, M., 3.5
Heagle, A. B., 2.5
Heap, B. R., 3.3
Hearmon, R. F. S., 4.1
Hearn, A. C., 1.4
Hecht, H. J., 2.3
Heel, M. van, 2.5
Hegerl, R., 2.5
Hehre, W. J., 1.2
Heideman, M. T., 1.3
Heil, P. D., 4.5
Heinermann, J. J. L., 2.2
Heiney, P. A., 4.4, 4.6
Helfrich, W., 4.4
Helgaker, T., 3.5
Helliwell, J. R., 2.2, 2.4
Hellner, E., 4.2
Helms, H. D., 1.3
Hende, J. van den, 1.3
Henderson, R., 2.3, 2.5
Hendricks, S., 4.2, 4.5
Hendrickson, W. A., 1.3, 2.2, 2.3, 2.4
Hendrikx, Y., 4.4
Hennion, B., 4.2
Hennion, M., 4.2
Henry, R., 3.5
Herglotz, G., 1.3
Herman, G. T., 2.5
Hermann, C., 1.3, 4.6
Hermans, J., 3.3
Herriot, J. R., 2.4
Herrmann, K. H., 2.5
Hewat, A., 3.3
Hewitt, J., 2.5
Heymann, J. A. W., 2.5
Higgs, H., 3.3
High, D. F., 2.3, 2.4
Hildebrandt, G., 5.1
Hills, G. J., 2.5
Hirabayashi, M., 2.5
Hiraga, K., 2.5
Hirsch, P. B., 2.5, 4.5, 5.1, 5.2
Hirschman, I. I. Jr, 1.3
Hirshfeld, F. L., 1.2, 2.3
Hirt, A., 2.5
Hirth, J. P., 4.4
Hitchcock, P. B., 4.4
Hjertén, S., 2.3
Hlinka, J., 4.2
Ho, M.-H., 2.5
Ho, M.-S., 2.5
Hockney, R. W., 3.5
Hodgkin, D. C., 2.2, 2.3, 2.4
Hodgson, K. O., 2.4
Hodgson, M. L., 1.3
Hoffmann, J.-U., 4.2
Hofmann, D., 4.5
Hofmann, D. W. M., 3.5
Hogle, J., 2.3, 3.3
Hohlwein, D., 4.2
Hohn, M., 2.5
Høier, R., 2.5, 4.3, 5.2
Hol, W. G. J., 2.3, 3.3
Holbrook, S. R., 1.3, 2.4
Hollenberg, J., 2.5
Holm, C., 3.5
Holmes, K. C., 2.5, 4.5
Honegger, A., 3.3
Hong, H., 4.2
Honjo, G., 4.3
Hopfinger, A. J., 3.3, 4.5
Hopgood, F. R. A., 3.3
Hoppe, W., 1.3, 2.2, 2.3, 2.4, 2.5
Horalik, L., 5.3
Horjales, E., 3.3
Hörmander, L., 1.3
Horn, P. M., 4.4
Horne, M. A., 5.3
Hornreich, R. M., 4.4
Hornstra, J., 2.3
Horstmann, M., 2.5
Hosemann, R., 2.3, 4.2, 4.5
Hoser, A., 4.2
Hoshino, S., 5.3
Hosoya, S., 5.3
Hosur, M. V., 2.3
Houston, T. E., 3.3
Hovmöller, S., 1.4, 2.5
Howe, J. M., 2.5, 4.2
Howells, E. R., 2.1
Howells, R. G., 2.2
Howie, A., 2.5, 4.3, 4.5, 5.2
Hradil, K., 4.2
Hrdlička, Z., 5.3
Hsiou, Y., 2.3
Hsiung, H., 4.4
Hu, Ch., 4.2
Hu, H., 4.5
Hu, H. H., 2.5
Huang, C., 3.3
Huang, C. C., 2.5, 4.4
Huang, K., 1.3, 4.1
Huang, Z., 2.5
Hubbard, R. E., 3.3
Huber, R., 1.3, 2.3, 2.4
Hudson, L., 4.5
Hudson, P. J., 3.3
Huesman, R. H., 2.5
Hughes, D. E., 1.3
Hughes, E. W., 1.3, 2.2, 2.3
Hughes, J. F., 3.3
Hughes, J. J., 2.3, 2.4
Hughes, S. H., 2.3
Hull, S., 4.2
Hull, S. E., 2.2
Hüller, A., 4.2
Hummel, W., 3.3
Hummelink, T., 3.3
Hummelink-Peters, B. G., 3.3
Hummer, G., 3.5
Hümmer, K., 5.1
Humphreys, C. J., 2.5, 4.3, 5.2
Hunsmann, N., 2.5
Hunt, J. F., 2.3
Huntingdon, H. B., 4.1
Hurley, A. C., 1.2, 2.5, 5.2
Huse, D. A., 4.4
Hutching, M. T., 4.2
Iannelli, P., 4.5
Ibers, J. A., 2.4, 4.2
Iijima, S., 4.3
Iizumi, M., 4.2
Ilag, L. L., 2.3
Imamov, R. M., 2.5
Immirzi, A., 1.3
Imry, Y., 4.4
Indenbom, V. L., 5.3
Ingram, V. M., 2.3, 2.4
Inoue, A., 2.5
Iolin, E. M., 5.3
Irwin, M. J., 2.2
Isaacs, N. W., 1.3, 2.2, 2.4, 3.3
Ishida, M., 2.5
Ishihara, K. N., 2.5, 4.6
Ishii, T., 4.2
Ishikawa, I., 5.3
Ishikawa, T., 5.1
Ishikawa, Y., 4.2
Ishizuka, K., 2.5
Isoda, S., 2.5, 4.5
Israel, R., 2.2
Ito, T., 3.2
Ivanova, M. I., 4.5
Ivantchev, S., 1.5
Iwata, H., 4.2
Izumi, F., 3.3

Jach, T., 5.1
Jack, A., 1.3, 2.4
Jackson, J. I., 2.5
Jacobson, R. A., 2.2, 2.3
Jacques, J., 4.4
Jaeger, J. C., 1.3
Jagodzinski, H., 4.2
Jahn, W., 2.5
Jakana, J., 2.5
James, R. W., 1.2, 1.3, 2.3, 4.2, 5.1
Jan, J.-P., 1.5
Janiak, M. J., 4.4
Janner, A., 1.3, 2.5, 4.2, 4.6
Janot, C., 4.2
Janot, Chr., 4.6
Jansen, L., 1.5
Janssen, P. A. J., 2.3
Janssen, T., 1.3, 1.5, 2.5, 4.2, 4.6
Jap, B. K., 2.5
Jaric, M. V., 4.2, 4.6
Jarić, M. Y., 2.5
Jarvis, L., 3.3
Jaynes, E. T., 1.3, 2.2
Jefferey, J. W., 4.2
Jeffery, B. A., 2.3
Jeffrey, G. A., 1.3
Jenni, S., 2.5
Jensen, L. H., 1.3, 2.3, 2.4
Jeu, W. H. de, 4.4
Jiang, J.-S., 2.3
Jia-xing, Y., 2.2
Jih, J. H., 2.3
Jogl, G., 2.3
Johannisen, H., 2.3
Johnson, A. W. S., 2.5, 5.2
Johnson, C. K., 1.2, 3.1, 3.3
Johnson, D. H., 1.3
Johnson, D. L., 4.4
Johnson, H. W., 1.3
Johnson, J. E., 2.3, 2.5, 3.3
Johnson, L. N., 2.3, 2.4, 3.3
Johnson, R. W., 1.3
Jolles, P., 2.3
Jones, B., 4.4
Jones, R., 4.2, 4.5
Jones, R. C., 4.2
Jones, T. A., 2.3, 2.5, 3.3
Jones, W., 3.5
Jones, Y., 2.3
Jorgensen, P., 3.5
Josefsson, T. W., 4.3
Joyeux, L., 2.5
Joyez, G., 4.2
Jürgensen, H., 1.3
Just, W., 4.2

AUTHOR INDEX

Kabsch, W., 3.3, 4.5
Kac, M., 1.3
Kaczmarz, S., 2.5
Kadečková, S., 5.3
Kaenel, R. A., 1.3
Kagan, Yu., 5.3
Kainuma, Y., 4.3, 5.2
Kaiser, H., 5.3
Kaiser-Bischoff, I., 4.2
Kakinoki, J., 4.2
Kalantar, A. H., 3.2
Kaldor, U., 2.1
Kalning, M., 4.2, 4.6
Kamada, K., 5.3
Kambe, K., 2.5, 5.2
Kamer, G., 2.3
Kamper, J., 2.3
Kaneko, K., 2.5
Kaneyama, T., 2.5
Kang, S. J., 3.3
Kannan, K. K., 2.3
Kansy, K., 3.3
Kaplan, D. R., 5.1
Kaplan, M., 4.4
Kara, M., 1.2
Karamertzanis, P. G., 3.5
Karbach, A., 4.5
Karle, I. L., 2.2
Karle, J., 1.3, 2.1, 2.2, 2.3, 2.4, 2.5
Kármán, T. von, 4.1
Karplus, M., 1.3, 3.3
Karrass, A., 1.3
Kartha, G., 2.3, 2.4
Kasper, J. S., 1.3, 2.2
Kasting, G. B., 4.4
Katagawa, T., 5.1
Katayama, K., 4.5
Kato, K., 4.6
Kato, N., 5.1, 5.3
Katsube, Y., 2.4
Katz, L., 3.3
Katznelson, Y., 1.3
Kaufman, M. J., 2.5
Kawaguchi, A., 4.5
Ke, E. Y., 2.5
Kearsley, S. K., 3.3
Keeling, J., 2.3
Keen, D. A., 4.2
Kek, S., 4.2, 4.6
Keller, A., 4.5
Keller, E., 3.3
Keller, J., 4.2
Kelley, B., 4.4
Kelly, B. A., 4.4
Kelton, K. F., 4.2, 4.6
Kendall, M., 1.2, 2.1
Kendrew, J. C., 1.3, 2.3, 2.4
Kennard, O., 3.1, 3.3
Kessel, M., 2.5
Kessler, M., 3.3
Ketelaar, J. A. A., 2.3
Khalak, H. G., 2.2
Khinchin, A. I., 1.3
Kiefer, J. E., 1.3, 2.1
Kiely, C. J., 2.5
Kikuta, S., 5.1, 5.3
Kim, M., 2.5
Kim, S.-H., 1.3, 2.3, 2.4
Kirby, I., 2.3
Kirichuk, V. S., 2.5
Kirkland, E. J., 2.5
Kirov, A., 1.4, 1.5
Kiselev, N. A., 2.5
Kissinger, C. R., 2.3
Kitagaku, M., 4.3
Kitagawa, Y., 2.4
Kitaigorodskii, A. J., 4.1
Kitaigorodsky, A. I., 4.2
Kitamura, N., 4.3
Kittel, C., 3.5, 4.4
Kitz, N., 1.3
Kjeldgaard, M., 2.5, 3.3
Klapperstück, M., 4.4
Klar, B., 5.3
Klechkovskaya, V. V., 2.5
Klei, H. E., 2.3
Klein, A. G., 5.3
Klein, M. L., 3.5
Kleinstück, K., 5.3
Kleman, M., 4.6
Klepeis, J. L., 3.5
Kleywegt, G. J., 2.3
Klimkovich, S., 1.4
Klug, A., 1.3, 2.2, 2.3, 2.5, 3.3, 4.5
Klug, H. P., 4.2
Kluyver, J. C., 1.3
Knol, K. S., 2.4
Knowles, J. W., 5.3
Kobayashi, S., 4.4
Kobayashi, T., 2.5
Koch, E., 1.1, 1.4
Koch, M. H. J., 2.2
Kodera, S., 4.3
Koellner, M., 5.3
Koetzle, T. F., 2.4, 3.3
Kogiso, M., 2.5, 5.2
Kohn, V. G., 5.1
Kohra, K., 5.1, 5.3
Kokkinidis, M., 2.3
Kolar, H., 2.5
Kolba, D. P., 1.3
Kolodziej, S., 2.5
Komada, T., 2.5
Komorek, M., 4.2
Komura, Y., 4.2
Kong, Y. F., 2.5
Konnert, J. H., 1.3, 2.4
Kopka, M. L., 2.3
Kopp, S., 2.5, 4.5
Kopperschlager, G., 2.5
Korekawa, M., 4.2, 4.6
Koritsansky, T., 3.5
Korn, D. G., 1.3
Körner, T. W., 3.5
Korpiun, P., 5.3
Kortan, A. R., 4.4
Kossel, W., 2.5
Koster, A. J., 2.5
Kosterlitz, J. M., 4.4
Kosykh, V. P., 2.5
Kovacs, A. J., 4.5
Kovalchuk, M. V., 5.1
Kovalev, O. V., 1.5
Koymans, L., 2.3
Krabbendam, H., 2.2
Krahl, D., 2.5, 4.3
Krämer, S., 2.5
Kraus, W., 3.3
Kraut, J., 1.3, 2.3, 2.4
Kress, W., 4.1
Kreuger, R. J., 5.2
Kriegman, D. J., 2.5
Krisch, M., 4.1
Krishna, P., 4.2
Krivanek, O. L., 4.3
Krivoglaz, M. A., 4.2, 4.3
Kroeker, M., 3.3
Kroon, J., 2.2, 3.5
Kroumova, E., 1.5
Kruse, F. H., 1.3
Kuchta, L., 2.2
Kühlbrandt, W., 2.5
Kühne, T., 2.5
Kuhs, W. F., 1.2
Kukla, D., 2.4
Kulda, J., 5.3
Kulidzhanov, F. G., 5.3
Kuligin, A. K., 2.5
Kulka, D., 1.3
Kull, F. J., 2.5
Kuntz, I. D., 3.3
Kuo, K. H., 2.5
Kurihara, K., 4.2
Kuriyan, J., 1.3
Kurki-Suonio, K., 1.2
Kuszewski, J., 2.3
Kutateladze, T. G., 2.5
Kutznetsov, P. I., 1.2
Kuvdaldin, B. V., 5.3
Kuwabara, S., 2.5
Kvardakov, V. V., 5.3
Kycia, S., 4.2
Kyriakidis, C. E., 2.2

Lacour, T. F. M., 1.3
Ladbrooke, B. D., 4.4
Ladisa, M., 2.2
Ladjadj, M., 2.5
Lagacé, L., 2.3
Lagendijk, R. L., 2.5
Lagomarsino, S., 5.1, 5.3
Lajzerowicz, J., 4.4
Lajzérowicz, J., 2.2
Laloe, F., 1.2
Lambert, D., 5.3
Lambert, M., 4.4
Lambert, M. A., 2.5
Lambiotte, J. J. Jr, 1.3
Lamfers, H.-J., 4.2
Lamy, J., 2.5
Lamzin, V. S., 2.3
Lancon, F., 4.6
Lanczos, C., 1.3
Landau, H. J., 1.3
Landau, L. D., 4.4
Lando, J. B., 4.5
Lang, A. R., 5.1
Lang, S., 1.3
Lang, W. W., 1.3
Langer, R., 2.5
Langridge, R., 3.3, 4.5
Langs, D. A., 2.2, 2.5, 4.5
Lanzavecchia, S., 2.5
Larine, M., 1.4
Larmor, J., 1.3
Lata, K. R., 2.5
Lattman, E. E., 2.3, 2.4
Laue, M., 1.1
Laue, M. von, 1.3, 5.1
Laurette, I., 2.5
Laval, J., 4.1
Laves, R., 4.2
Lavoine, J., 1.3
Lawrence, M. C., 2.5
Lawrence, P. D., 2.5
Lawson, K. D., 4.4
Lawton, J. A., 2.5
Le Bail, A., 3.3
Le Guillou, J. C., 4.4
Lea, S. M., 2.3
Leadbetter, A. J., 4.4
Leapman, R. D., 4.3
Lechner, R. E., 4.2
Lederer, F., 3.3
Ledermann, W., 1.3, 4.1
Lee, E. J., 4.5
Lee, H., 3.5
Lee, J. C., 2.5
Lee, P. L., 4.2
Lee, S. D., 4.4
Lee, Y.-R., 5.1
Leenhouts, J. I., 2.3
Lefebvre, J., 4.1, 4.2
Lefebvre, S., 4.2
Lefeld-Sosnowska, M., 5.1
Legg, M. J., 1.3
Lehmann, M., 5.3
Lehmann, M. S., 1.3, 2.2
Lehmpfuhl, G., 2.5, 4.3
Lei, J., 4.2
Leiman, P. G., 2.3
Leith, A., 2.5
Lele, S., 4.2
Lemoine, G., 3.3
Lentz, P. J. Jr, 2.3
Lepault, J., 2.5
Lerch, M., 4.2
Lerner, F. Ya., 2.5
Lescar, J., 5.1
Lescoute, A., 2.5
Lesk, A. M., 3.3
Leslie, A. G. W., 1.3, 2.3, 3.3
Lessinger, L., 2.2
Leszczynski, M., 4.1
Leung, P., 1.2
Leusen, F. J. J., 3.5
Leusen, F. J. L., 3.5
Levanyuk, A. P., 4.4
Levelut, A. M., 4.4
Levens, S. A., 1.3
Levine, D., 2.5, 4.6
Levinthal, C., 3.3
Levitov, L. S., 4.6
Levitt, M., 1.3, 2.4, 3.3
Levy, H. A., 1.2, 1.3, 3.1
Lewis, M., 2.3
Lewis, T. C., 3.5
Lewitt, R. M., 2.5
Li, D. X., 2.5
Li, F. H., 2.5
Li, J. Q., 2.5
Li, Y., 2.5
Liang, C., 3.5
Liang, K. S., 4.4
Liang, Y. Y., 2.5
Liebert, L., 4.4
Liebert, L. E., 4.4
Liebman, G., 2.5
Lien, S. C., 4.4
Lieth, C. W. van der, 3.3
Lievert, L., 4.4
Lifchitz, A., 1.3, 2.3
Lifshitz, E. M., 4.4
Lifson, S., 3.3
Lighthill, M. J., 1.3
Lijk, L. J., 3.3
Liljas, L., 2.3, 3.3
Liljefors, T., 3.3
Linares-Galvez, J., 5.3
Lindegaard, A., 4.4
Lindsey, J., 2.3
Link, V., 4.4
Linnik, I. Ju., 1.3
Lipanov, A. A., 4.5
Lipkowitz, K. B., 3.3
Lippert, B., 2.1
Lipscomb, W. N., 2.3
Lipson, H., 1.1, 1.2, 1.3, 1.4, 2.1, 2.3, 4.2, 4.5
Litster, J. D., 4.4
Litvin, D. B., 2.3
Liu, J., 4.5
Liu, W., 2.5
Liu, Y.-W., 2.5
Livanova, N. B., 2.5
Livesey, A. K., 1.3
Lloyd, T. E., 2.5
Loane, R. F., 4.3
Lobachev, A. N., 2.5
Lobanova, G. M., 2.5
Lobert, S., 4.5
Lock, C. J. L., 2.1
Lockhart, T. E., 4.4
Lomer, T. R., 2.1
Lomont, J. S., 1.5
Lonsdale, K., 1.3
Lontovitch, M., 5.2
Looijenga-Vos, A., 4.6
Lorenz, M., 4.5
Loris, R., 5.1
Lotz, B., 2.5, 4.5
Love, W., 1.3
Love, W. E., 2.2, 2.3
Love, W. F., 1.5
Lovell, F. M., 1.3
Lövgren, S., 2.3
Lowde, R. D., 5.3
Lu, C., 1.3
Lu, G., 2.3
Luban, M., 4.4
Lubensky, T. C., 4.4
Lucas, B. W., 4.2
Luck, J. M., 4.6
Ludewig, J., 5.1
Ludtke, S. C., 2.5

AUTHOR INDEX

Ludtke, S. J., 2.5
Luenberger, D. G., 3.3
Luić, M., 2.2
Lunin, V. Y., 2.3
Lunin, V. Yu., 1.3
Lunina, N. L., 2.3
Luo, M., 2.3
Lurie, N. A., 4.1
Lurz, R., 2.5
Lushington, K. J., 4.4
Luther, P., 4.5
Luty, T., 4.1
Luzzati, V., 2.3, 4.4
Lybanon, M., 3.2
Lyman, P. F., 5.1
Lynch, D. F., 2.5, 5.2
Lynch, R. E., 2.3

Ma, J. P., 2.5
Ma, Q., 2.3
Ma, S. K., 4.4
MacGillavry, C. H., 1.3, 4.5
Machin, P. A., 3.3
Mackay, A. L., 2.2, 2.5, 3.3
MacKay, M., 2.3
MacLane, S., 1.3
MacNicol, D. D., 2.5
Macovski, A., 2.5
Macrae, C. F., 3.3
MacRae, T. P., 4.5
Mada, H., 4.4
Madariaga, G., 1.5
Madden, P. A., 3.5
Madelung, E., 3.4
Madhav Rao, L., 5.3
Magdoff, B. S., 2.3
Magnus, W., 1.3
Mahendrasingam, A., 4.5
Mahon, M., 3.3
Maier, W., 4.4
Main, P., 1.3, 2.2, 2.3, 2.5
Makowski, L., 4.5
Malgrange, C., 5.1, 5.3
Malik, K. M. A., 4.4
Maling, G. C., 1.3
Malladi, R., 2.5
Mallick, S. P., 2.5
Mallikarjunan, M., 3.3
Malthête, J., 4.4
Maly, K., 4.2, 4.6
Mandelkern, L., 4.5
Mandelkow, E., 4.5
Mani, N. V., 2.4
Manley, R. St. J., 2.5
Mannami, M., 2.5
Mao, Y., 2.5
Marabini, R., 2.5
Marchington, B., 1.3
Mardix, S., 4.2
Marel, R. P. van der, 2.1
Marigo, A., 4.2
Marinder, B. O., 2.5
Mark, H., 2.4
Markham, R., 2.5
Marko, M., 2.5
Marks, L. D., 4.3
Marsh, R. E., 3.2, 3.5
Marson, F., 4.4
Martin, C., 2.3
Martin, P. C., 4.4
Martinez-Miranda, L. J., 4.4
Martorana, A., 4.2
Marumo, F., 1.2
Marvin, D. A., 1.3, 4.5
Marynissen, H., 4.4
Masaki, N., 5.3
Maslen, V. W., 1.2
Maslen, W. V., 4.3
Mason, R., 4.4
Mason, S. A., 4.5
Massariol, M.-J., 2.3
Massidda, V., 3.4
Mastryukov, V. S., 2.5
Masumoto, K., 2.5
Masumoto, T., 2.5
Matadeen, R., 2.5
Matej, S., 2.5
Materlik, G., 1.2, 5.1
Mathews, F. S., 3.3
Mathiesen, R. H., 5.1
Mathiesen, S., 4.4
Mathieson, A. McL., 2.3
Matsubara, E., 4.2
Matsuda, T., 2.5
Matthews, B. W., 1.3, 2.3, 2.4
Mauguen, Y., 2.2, 2.4
Mauritz, K. A., 4.5
Max, N. L., 3.3
Mayer, J., 2.5
Mayer, S. W., 1.3
Mayers, D. F., 4.3
Mazeau, K., 4.5
Mazid, M. A., 4.4
Mazkedian, S., 5.3
Mazuré-Espejo, C., 5.3
Mazzarella, L., 2.3, 2.4
McArdle, P., 3.3
McCabe, P., 3.3
McCall, M. J., 3.3
McClellan, J. H., 1.3
McCourt, M. P., 2.5, 4.5
McCoy, A. J., 2.3
McDonald, W. S., 2.4
Mcewen, B., 2.5
McFarland, K., 3.5
McGreevy, R. L., 4.2
McIntyre, G. J., 1.2
McKean, H. P., 1.3
McKenna, R., 2.3
Mckernen, S., 2.5
McLachlan, A. D., 3.3
McLachlan, D., 2.3, 2.5
McMahon, B., 1.4
McMillan, W. L., 4.4
McMullan, R. K., 3.3
McMurchie, L. E., 3.5
McPherson, A., 2.4
McQueen, J. E., 3.3
McWhan, D. B., 4.4
Mechin, I., 2.5
Meiboom, S., 4.4
Meichle, M., 4.4
Melone, S., 5.3
Mendiratta, S. K., 5.3
Meng, E. C., 2.5
Menzer, G., 2.3
Mermin, N. D., 1.1, 4.6
Mersereau, R. M., 1.3, 2.5
Merwe, J. H. van der, 4.4
Messiah, A., 5.2
Metropolis, N., 4.2
Meyer, C. E., 4.3
Meyer, C. H., 2.5
Meyer, E. F., 3.3
Meyer, G., 2.5
Meyer, L. B., 2.5
Meyer, R. B., 4.4
Michalec, R., 5.3
Michejda, C. J., 2.3
Micu, A. M., 4.2
Midgley, P. A., 2.5
Mielke, T., 2.5
Mierzejewski, A., 4.1
Mighell, A. D., 2.3
Mikula, P., 5.3
Millane, R. P., 4.5
Miller, A., 4.5
Miller, D. P., 4.5
Miller, G. H., 4.5
Miller, J. R., 3.3
Miller, J. S., 4.2
Miller, R., 2.2, 4.5
Miller, S. C., 1.5
Miller, S. T., 2.3
Mills, D. M., 5.1
Mimori-Kiyosue, Y., 4.5
Minakawa, N., 5.3
Mindell, J. A., 2.5
Ming, D. M., 2.5
Mitra, A. K., 4.5
Mitra, K., 2.5
Mitsui, T., 4.5
Miyake, S., 2.5
Miyano, K., 4.4
Miyazaki, M., 2.5
Mo, F., 5.1
Mo, Y. D., 2.5
Moereels, H., 2.3
Moliere, G., 4.3
Moliterni, A. G. G., 2.2
Möllenstedt, G., 2.5
Moncrief, J. W., 2.3
Moncton, D. E., 4.4
Montroll, E. W., 1.3
Moodie, A. F., 2.5, 5.2
Mooji, W. T. M., 3.5
Moon, P. B., 2.4
Mooney, P. E., 4.3
Moore, D. H., 1.3
Moore, P. B., 2.5
Moras, D., 2.3
More, M., 4.2
Morffew, A. J., 3.3
Morgenroth, W., 4.2
Mori, M., 4.3
Moriguchi, S., 2.5
Morimoto, C. N., 3.3
Morinaga, M., 4.2
Moring, I., 2.3
Moritz, W., 4.2
Morniroli, J. P., 2.5
Morris, E. P., 2.5
Morris, J., 3.5
Morris, R. L., 1.3
Moser, W. O. J., 1.3
Mosley, A., 4.4
Moss, B., 2.5
Moss, D. S., 4.2
Moss, G., 1.2
Moss, S. C., 4.2, 4.3
Mosser, A. G., 2.3
Motherwell, W. D. S., 3.3, 3.5
Motohashi, H., 5.3
Mouche, F., 2.5
Moussa, F., 4.4
Moustiakimov, M., 2.2
Muirhead, H., 2.3, 2.4
Mukamel, D., 4.4
Mukherjee, A. K., 2.2
Mullapudi, S., 2.5
Müller, H., 4.2
Müller, R., 5.1
Müller, U., 4.2
Munn, R. W., 3.4
Murakami, W. T., 2.3
Murdock, W. L., 1.3
Murray, W., 3.3
Murshudov, G. N., 2.3
Murthy, M. R. N., 2.3
Muus, I. T., 4.5
Myller-Lebedeff, W., 2.1

Nagabhushana, C., 4.4
Nagasawa, T., 2.5, 5.2
Nagem, R. A. P., 2.3
Naiki, T., 2.5
Nakatsu, K., 2.3
Namba, K., 4.5
Nambudripad, R., 4.5
Narayan, R., 1.3, 2.2, 2.4
Narayanan, B. A., 3.5
Natarajan, P., 2.5
Nathans, R., 4.2
Natterer, F., 1.3, 2.5
Navaza, J., 1.3, 2.2, 2.3
Nave, C., 1.3, 4.5
Navia, M. A., 2.4
Nawab, H., 1.3
Naya, S., 2.2, 4.5
Neder, R. B., 4.2
Neisser, J. Z., 4.5
Nelson, D. E., 1.3
Nelson, D. R., 4.2, 4.4
Nelson, H. M., 1.5
Neto, A. M. F., 4.4
Neubert, M. E., 4.4
Neubüser, J., 1.3
Newham, R. J., 3.4
Newman, W. M., 3.3
Newsam, J. M., 3.5
Neyertz, S., 3.5
Ng, E. G., 2.5
Niall, H. D., 3.3
Nicastro, D., 2.5
Nicholson, P. B., 4.5
Nicholson, R. B., 2.5, 5.2
Nickell, S., 2.5
Nickitenko, A., 2.5
Nieh, Y.-P., 2.3
Nield, V. M., 4.2
Nierhaus, K. H., 2.5
Nigam, G. D., 2.1
Niggli, A., 1.3
Niimura, N., 4.2
Nijboer, B. R. A., 3.4, 3.5
Nilges, M., 2.3
Nishimura, D. G., 2.5
Nissen, P., 2.5
Nitsch, M., 2.5
Nitta, I., 2.2
Nityananda, R., 1.3, 2.2
Nixon, P. E., 2.3
Nolze, G., 3.3
Nonoyama, M., 4.3
Nordman, C. E., 1.3, 2.2, 2.3
North, A. C. T., 2.3, 2.4, 3.3
Norton, D. A., 2.2
Nosé, S., 3.5
Nowacki, W., 2.5
Nowell, H., 3.5
Nunzi, A., 2.2
Nussbaumer, H. J., 1.3

Öberg, B., 2.3
Oberhettinger, F., 1.3
Oberteuffer, J. A., 5.3
Oberti, R., 2.4
Ocko, B. M., 4.4
Oda, T., 2.2
O'Donnell, T. J., 3.3
Oesterhelt, D., 2.5
Ogata, Y., 2.5
Ogawa, T., 2.5
Ogburn, K. D., 2.5
Ohara, M., 4.5
Ohshima, K., 4.2, 4.3
Ohtsuki, Y. H., 4.3, 5.1
Oikawa, T., 4.3
Okabe, A., 2.5
Okaya, J., 2.2
Okaya, Y., 2.2, 2.3, 2.4
O'Keefe, M. A., 2.5
Olafson, B. D., 3.3
Olmer, P., 4.1
Olsen, J., 3.5
Olsen, K. W., 2.3
Olson, A. J., 1.3, 2.3, 3.3
Olthof-Hazekamp, R., 1.4, 2.2
Omura, T., 4.3
Ono, A., 2.5
Onsager, L., 1.3, 4.4
Opat, G. I., 5.3
Opdenbosch, N. van, 3.3
Ord, K., 2.1
Orlando, R., 3.5
Orlov, S. S., 2.5
Orlova, E. V., 2.5
Ørmen, P.-J., 1.2
Ostermann, A., 4.2
O'Sullivan, J. D., 2.5
Oszlányi, G., 2.2
Ott, H., 2.2
Ottensmeyer, F. P., 2.5
Overduin, M., 2.5
Overhauser, A. W., 4.2, 5.3
Ozawa, T. C., 3.3

AUTHOR INDEX

Pabst, M., 4.1
Paciorek, W. A., 4.6
Pähler, A., 2.2
Paley, R. E. A. C., 1.3
Palleschi, V., 4.4
Palmer, M. R., 2.5
Palmer, R. A., 2.3, 2.4
Palmer, S. B., 5.3
Pan, M., 2.5
Pan, Q., 2.5
Pandey, D., 4.2
Panepucci, E. H., 2.3
Pannu, N. S., 2.3
Pantelides, C. C., 3.5
Paoletti, A., 4.2
Pape, T., 2.5
Paradossi, G., 4.5
Park, H., 4.5
Parks, T. W., 1.3
Parlett, B. N., 2.5
Parmon, V. S., 2.5
Parodi, O., 4.4
Parsey, J. M. Jr, 4.2
Parthé, E., 4.3
Parthasarathy, R., 2.3, 2.4
Parthasarathy, S., 2.1, 2.2, 2.4
Pascual-Montano, A., 2.5
Pashley, D. W., 2.5, 4.5, 5.2
Pastore, A., 3.3
Patel, J. R., 5.1
Patel, K., 3.5
Pattabiraman, N., 3.3
Pattanayek, R., 4.5
Patterson, A. L., 1.1, 1.3, 2.3, 2.4, 4.2
Patterson, C., 5.3
Paturle, A., 1.2
Patwardhan, A., 2.5
Pätzold, H., 4.3
Paul, D., 2.5
Pauling, L., 1.3, 2.3
Paulmann, C., 4.2
Pauwels, R., 2.3
Pavelcík, F., 2.2
Pavlovitch, A., 4.6
Pavone, P., 4.1
Pawley, G. S., 4.1
Pearce, L. J., 3.3
Pearl, L. H., 3.3
Pearlman, D. A., 3.5
Pearson, J., 3.3
Pearson, K., 1.3
Pease, M. C., 1.3
Pedersen, B., 4.2
Pedersen, L., 3.5
Peerdeman, A. F., 2.2, 2.3, 2.4
Peierls, R. E., 4.4
Peisl, J., 4.2
Penczek, P. A., 2.5
Penning, P., 5.1
Penrose, R., 2.5, 4.6
Penzkofer, B., 4.2
Pepinsky, R., 1.3, 2.2, 2.3, 2.4
Perera, L., 3.5
Perez, S., 4.5
Pérez, S., 2.5
Perez-Mato, J. M., 1.4, 1.5, 4.2
Perham, R. N., 4.5
Perrakis, A., 2.3
Perram, J. W., 3.5
Perrier de la Bathie, R., 5.3
Pershan, P. S., 4.4
Perutz, M. F., 2.2, 2.3, 2.4
Peschar, R., 2.2
Petef, G., 2.3
Peters, C., 1.3
Petersen, H. G., 3.5
Petrascheck, D., 5.3
Petricek, V., 4.2, 4.6
Pétroff, J. F., 5.3
Petrov, V. V., 1.3
Petrova, T. E., 2.3
Petržílka, V., 5.3
Petsko, G. A., 1.3, 3.3
Pettersen, E. F., 2.5
Pezerat, H., 4.2
Pfaff, G., 3.3
Pflanz, S., 4.2
Pflugrath, J. W., 2.3
Phillips, D. C., 2.1, 2.2, 2.3, 2.4, 3.3
Phillips, J. C., 2.4
Phillips, S. E. V., 3.3
Phizackerley, R. P., 2.4
Phong, B. T., 3.3
Pickworth, J., 2.3
Pielartzik, H., 4.5
Pietila, L.-O., 3.4
Pietronero, L., 4.2
Pietsch, U., 2.5
Pifferi, A., 2.2
Pigram, W. J., 4.5
Pillardy, J., 3.5
Pilling, D. E., 1.3
Pindak, R., 4.4
Pink, M. G., 3.3
Pinsker, Z. G., 2.5, 5.1
Piquemal, J. P., 3.5
Pirie, J. D., 4.1
Piro, O. E., 2.2
Plano, R. J., 4.4
Plotnikov, A. P., 2.5
Plotnikov, V. P., 2.5
Pochon, F., 2.5
Podjarny, A. D., 2.2, 2.3, 2.4
Podurets, K. M., 5.3
Pogany, A. P., 2.5
Pokrovsky, V. L., 4.4
Polder, D., 5.1
Polidori, G., 2.2
Polikarpov, I., 2.3
Poljak, R. J., 2.3
Pollack, H. O., 1.3
Pollock, E. L., 3.5
Ponder, J. W., 3.5
Poole, C. P., 3.3
Popa, N. C., 4.1
Pople, J. A., 1.2
Popp, D., 4.5
Porter, T. K., 3.3
Porter, W., 3.5
Portier, R., 2.5, 5.2
Potenzone, R., 3.3
Potter, C. S., 2.5
Potterton, E. A., 3.3
Potts, R. B., 1.3
Pouget, J. P., 4.2
Powell, B. M., 4.2
Powell, M. J. D., 2.5
Prandl, W., 4.2
Prange, T., 2.3
Prasad, B. V. V., 2.5
Pratt, L. R., 3.5
Press, W., 4.2, 4.6
Preston, A. R., 2.5
Price, L. S., 3.5
Price, S. L., 3.5
Prick, A. J., 2.2
Prins, J. A., 2.4, 4.2, 5.1
Proffen, Th., 4.2
Prosen, R. J., 1.3, 2.3
Prost, J., 4.4
Prout, C. K., 3.3
Provencher, S. W., 2.5
Pryor, A. W., 4.1, 4.6
Pulay, P., 3.5
Puliti, P., 5.3
Pullan, L., 2.5
Purisima, E. O., 3.3
Pustovskikh, A. I., 2.5
Pusztai, L., 4.2
Pynn, R., 4.1

Qian, C., 2.3
Quandalle, P., 1.3
Quick, J., 3.5
Quilichini, M., 4.2
Quiocho, F. A., 2.5, 3.3
Qurashi, M. M., 1.3
Rabinovich, D., 2.1, 2.3
Rabinovich, S., 2.1
Rabson, D. A., 4.6
Rackham, G. M., 2.5
Rader, C. M., 1.3
Radermacher, M., 2.5
Radha, A., 4.5
Radhakrishnan, R., 3.3
Radi, G., 2.5
Radons, W., 4.2
Rae, A. D., 2.2, 2.3
Raghavacharyulu, I. V. V., 1.5
Raghavan, N. V., 2.4
Raghavan, R. S., 2.4
Rahman, S. H., 4.2
Raimondi, D. L., 1.2
Raïtman, E. A., 5.3
Raiz, V. Sh., 2.4
Raja, V. N., 4.4
Rajagopal, H., 2.4
Rajashankar, K. R., 2.3
Ramachandran, G. N., 2.2, 2.3, 2.4, 5.1
Ramagopal, U. A., 2.3
Raman, C. V., 4.1
Raman, S., 2.2, 2.3, 2.4
Ramaseshan, S., 2.3, 2.4
Ramaswamy, S., 4.4
Rango, C. de, 2.2, 2.4
Rao, R. R., 1.3
Rao, S. N., 2.3
Rao, S. T., 3.3
Raselli, A., 3.3
Rasmussen, B., 5.1
Rasmussen, K., 3.4
Ratna, B. R., 4.4
Rauch, H., 5.3
Raum, K., 5.3
Ravelli, R., 2.2
Rawiso, M., 4.4
Rayleigh (J. W. Strutt), Lord, 1.3, 2.1
Rayment, I., 2.3, 3.3
Read, R. J., 2.3
Redlack, A., 3.5
Rees, A. L. G., 2.5
Rees, D. C., 2.3
Refaat, L. S., 2.3
Reid, T. J. III, 2.3
Reif, F., 1.3
Reijen, L. L. van, 1.3
Reiner, I., 1.3
Reiss-Husson, F., 4.4
Reman, F. C., 4.4
Remillard, B., 4.5
Ren, J., 2.3
Ren, P. Y., 3.5
Renka, R., 2.5
Renninger, M., 5.1
Reuber, E., 2.5
Revol, J. F., 2.5
Rez, P., 2.5, 4.3, 5.2
Rhyner, J., 4.6
Ricci, R., 4.2
Rice, L. M., 2.3
Rice, S. O., 1.3
Richardson, J. S., 3.3
Richardson, J. W., 2.2
Richardson, R. M., 4.4
Rickert, S. E., 4.5
Riddle, A. C., 2.5
Riekel, C., 4.2
Riesz, M., 1.3
Rietveld, H. M., 4.2
Riley, D., 3.5
Rimmer, B., 2.3
Rini, J. M., 2.3
Rivard, G. E., 1.3
Rixon, F. J., 2.5
Robertson, J. H., 2.3
Robertson, J. M., 1.3, 2.3, 2.4, 3.2
Robinson, G., 1.3, 3.2
Rodewald, M., 4.3
Rodgers, J. R., 3.3
Rodgers, J. W., 2.4
Rodrigues, A. R. D., 5.3
Rodriguez-Carvajal, J., 3.3
Roetti, C., 1.2, 3.5
Rogers, D., 2.1, 2.2, 2.3
Rokhsar, D. S., 4.6
Rollett, J. S., 1.3, 3.3
Roseman, A. M., 2.5
Rosen, J., 1.5
Rosenbluth, A. W., 4.2
Rosenbluth, M. N., 4.2
Rosenstein, R. D., 2.3
Ross, C., 2.3
Rosshirt, E., 4.2
Rossmann, M. G., 1.3, 2.2, 2.3, 2.4, 3.3
Rossouw, C. J., 2.5, 4.3
Rouiller, I., 2.5
Roux, D., 4.4
Roversi, P., 2.3, 3.5
Rowlands, D., 2.3
Rowlands, R. J., 4.5
Rozenfeld, A., 2.5
Rueckert, R. R., 2.3
Ruedenberg, K., 1.2
Ruf, T., 4.1
Rugman, M., 4.4
Rühle, M., 2.5
Ruijgrok, Th. W., 3.5
Ruiz, T., 2.5
Ruland, W., 4.2
Rust, H.-P., 2.5
Rustichelli, F., 5.3
Ruston, W. R., 4.2
Rybnikar, F., 4.5
Rydén, L., 2.3
Rypniewski, W., 2.5
Ryskin, A. I., 2.5

Saad, A., 2.5
Sabine, T. M., 4.2
Sackmann, H., 4.4
Sadashiva, B. K., 4.4
Sadoc, J. F., 4.4
Sadova, N. I., 2.5
Safinya, C. R., 4.4
Safran, S. A., 4.4
Sagui, C., 3.5
Sahni, V. C., 4.1
Saibil, H. R., 2.5
Saito, M., 2.5
Saito, P., 2.5
Saito, R., 2.5
Saito, Y., 2.2, 2.3
Saitoh, K., 2.5
Saka, T., 5.1
Sakabe, K., 2.2
Sakabe, N., 2.2
Sakurai, K., 4.2
Salamon, M. B., 4.2
Sande, G., 1.3
Sander, B., 2.5, 5.3
Sandonis, J., 5.3
Sands, D. E., 1.1, 3.1
Sanjurjo, J. R., 2.5
Sansom, C., 4.2
Sarikaya, M., 2.5
Sarko, A., 4.5
Sasada, Y., 2.3, 3.3
Sato, H., 4.2
Satow, Y., 2.4
Saunders, M., 2.5
Saunders, V. R., 3.5
Saupe, A., 4.4
Sauvage, M., 4.3, 5.3
Saxton, W. O., 2.5
Sayre, D., 1.3, 2.2, 2.3, 2.4, 2.5, 4.5
Scaringe, P. R., 4.2
Scaringe, R. P., 2.5
Scatturin, V., 2.5
Schacher, G. E., 3.4
Schaetzing, R., 4.4
Schaffitzel, C., 2.5
Schärpf, O., 4.2
Schatz, M., 2.5
Schenk, H., 2.2
Scheraga, H. A., 3.3, 3.5

Scheres, S. H. W., 2.5
Scheringer, C., 1.2
Scherm, R., 5.3
Scherzer, O., 2.5
Schevitz, R. W., 2.2, 2.3, 2.4
Schilling, J. W., 2.3
Schiltz, M., 2.3
Schiske, P., 2.5
Schlenker, M., 5.3
Schmatz, W., 4.2
Schmidt, H. H., 5.3
Schmidt, M. U., 3.5
Schmidt, R., 2.5
Schmidt, T., 2.3
Schmidt, W. C. Jr, 3.3
Schnabel, W., 4.2
Schneider, A. I., 4.5
Schneider, T. R., 2.2, 2.3
Schoenberg, I. J., 3.5
Schoenborn, B. P., 2.4
Schofield, P., 4.1
Schomaker, V., 1.1, 1.2, 1.3, 2.3, 2.5, 3.2
Schomberg, H., 2.5
Schoone, J. C., 2.4
Schrader, H., 4.2
Schramm, H. J., 2.5
Schröder, M., 3.3
Schröder, R., 2.5
Schroder, R. R., 2.5
Schroeder, M. R., 1.3
Schroer, B., 2.5
Schuessler, H. W., 1.3
Schüler, M., 2.5
Schuller, D. J., 2.3
Schulz, H., 1.2, 4.2
Schulze, G. E. R., 5.3
Schumacker, R. A., 3.3
Schuster, S. L., 4.1
Schutt, C. E., 1.3, 2.3
Schwager, P., 1.3, 2.4
Schwartz, L., 1.3
Schwartz, L. H., 4.2
Schwarzenbach, D., 1.2
Schwarzenberger, R. L. E., 1.3
Schweika, W., 4.2
Schweizer, B., 3.5
Schweizer, J., 5.3
Scott, W. R., 1.3
Scudder, M. L., 2.4
Sears, V. F., 4.2, 5.3
Secomb, T. W., 2.5
Sedláková, L., 5.3
Seidl, E., 5.3
Seitz, E., 4.2
Seitz, F., 1.4
Sekii, H., 2.5, 5.2
Selinger, J. V., 4.4
Sellar, J. R., 5.2
Semiletov, S. A., 2.5
Senechal, M., 4.6
Serrano, J., 4.1
Sethna, J. P., 4.4
Sezan, M. I., 2.5
Sha, B. D., 2.5
Shaffer, P. A. Jr, 1.3
Shaikh, T., 2.5
Shakked, Z., 2.1, 2.3
Shan, N., 3.5
Shan, Y. B., 3.5
Shankland, K., 2.5, 4.5
Shannon, C. E., 1.3, 2.5
Shannon, M. D., 2.5
Shao-Hui, Z., 2.2
Shapiro, A., 3.3
Shappell, M. D., 2.3
Shashidhar, R., 4.4
Shashua, R., 2.1
Shastri, S. D., 5.1
Shaw, D. E., 3.5
Sheat, S., 2.3
Shechtman, D., 2.5, 4.6
Sheldrick, G. M., 2.2, 2.3
Shen, Q., 5.1
Shen, Y. R., 4.4
Shenefelt, M., 1.3
Sheriff, S., 2.3, 2.4
Sherry, B., 2.3
Sherwood, J. N., 4.2
Shilov, G. E., 1.3
Shil'shtein, S. Sh., 5.3
Shimanouchi, T., 3.3
Shipley, G. G., 4.4
Shirane, G., 4.1, 4.2
Shmueli, U., 1.3, 1.4, 2.1, 3.1, 3.2
Shoemaker, C. B., 2.3
Shoemaker, D. P., 1.3, 2.3
Shoemaker, V., 2.5
Shohat, J. A., 1.3
Shore, V. C., 2.4
Shortley, G. H., 1.2
Shotton, M. W., 4.5
Shtrikman, S., 4.4
Shull, C. G., 5.3
Sicignano, A., 2.3
Siddons, D. P., 5.3
Sidorenko, S. V., 2.5
Sieber, W., 3.3
Siegel, B. M., 2.5
Siegrist, T., 3.3
Sieker, L. C., 2.4
Sigaud, G., 4.4
Sigler, P. B., 2.2, 2.3, 2.4
Sigworth, F. J., 2.5
Sikka, S. K., 2.4
Silcox, J., 4.3
Siliqi, D., 2.2
Silverman, H. F., 1.3
Sim, G. A., 2.2, 2.3, 4.5
Simerska, M., 2.2
Simonov, V. I., 2.2, 2.3
Simonson, T., 2.3
Simpson, A. A., 2.3
Simpson, P. G., 2.3
Singh, A. K., 2.3, 2.4
Singleton, R. C., 1.3
Sinha, S. K., 3.5, 4.4
Sint, L., 2.2
Sippel, D., 5.3
Siripitayananon, J., 4.2
Sirota, E. B., 4.4
Sirota, M. I., 2.5
Sivardière, J., 5.3
Sivý, J., 2.2
Sixma, T. K., 2.3
Sjögren, A., 2.5
Skehel, J. J., 2.3
Skilling, J., 1.3
Skoglund, U., 2.3
Skoulios, A., 4.4
Skovoroda, T. P., 2.3
Skuratovskii, I. Y., 4.5
Slater, L. S., 1.5
Sleight, M. E., 2.5
Slepian, D., 1.3
Sluckin, T. J., 4.4
Sly, W. G., 1.3
Smaalen, S. van, 4.2, 4.6
Small, D., 4.4
Smit, B., 3.5
Smith, A. B. III, 4.4
Smith, D. J., 2.5
Smith, E. R., 3.5
Smith, G., 3.3
Smith, G. D., 2.2
Smith, G. S., 4.4
Smith, G. W., 4.4
Smith, J. C., 4.2
Smith, J. L., 2.2, 2.3
Smith, J. V., 1.5
Smith, M. B. K., 2.3
Smith, R. H. Jr, 2.3
Smith, T., 4.1
Smith, W., 3.5
Smits, J. M. M., 2.2
Smoluchowski, R., 1.5
Sneddon, I. N., 1.3
Soboleva, A. F., 2.5
Socolar, J. E. S., 4.6
Soeter, N. M., 2.3
Sokol'skii, D. V., 5.3
Soldani, M., 3.5
Solitar, D., 1.3
Solmon, D. C., 2.5
Solomon, L., 4.4
Somenkov, V. A., 5.3
Somers, D., 2.3
Sondhauss, P., 5.1
Soni, R. P., 1.3
Sorensen, L. B., 4.4
Sorzano, C. O. S., 2.5
Soyer, A., 3.3
Spackman, M. A., 3.5
Spagna, R., 2.2
Spahn, C. M. T., 2.5
Spanton, S., 3.5
Sparks, C. J., 1.2, 4.2
Sparks, R. A., 1.3, 3.3
Speake, T. C., 1.3
Speakman, J. C., 2.3
Spek, A. L., 2.2, 3.3
Spellward, P., 2.5
Spence, J. C. H., 2.5, 4.3
Spiegel, M. R., 2.1
Spink, J. A., 2.5
Spiwek, H., 3.5
Sprang, S. R., 3.3
Sprecher, D. A., 1.3
Springer, T., 4.2, 4.4
Sprokel, G. E., 4.4
Sproull, R. F., 3.3
Squire, J. M., 2.5, 4.5
Squires, G. L., 1.2, 4.1, 5.3
Srinivasan, R., 2.1, 2.2, 2.4
Staden, R., 1.3, 2.3, 3.3
Stahl, S. J., 2.5
Stammers, D., 2.3
Stanley, E., 2.2, 4.5
Stark, H., 2.5
Stark, W., 4.5
Stasiak, A., 2.5
Stassis, C., 5.3
States, D. J., 3.3
Staudenmann, J. L., 5.3
Stauffacher, C. V., 2.3
Steeds, J. W., 2.5
Stegemeyer, H., 4.4
Steger, W., 4.2
Stegun, I. A., 2.1
Steigemann, W., 1.3, 2.4
Stein, Z., 2.1
Steinberger, I. T., 4.2
Steinhardt, P. J., 2.5, 4.6
Steinkilberg, M., 2.5
Steinrauf, L. K., 2.3
Steitz, T. A., 2.3, 2.5
Stence, C. N., 2.3
Stephanik, H., 5.1
Stephen, M. J., 4.4
Stephens, P. W., 4.2, 4.4, 4.6
Stephenson, G. B., 4.4
Stephenson, R., 3.3
Stetsko, Yu. P., 5.1
Steurer, W., 2.5, 4.2, 4.6
Steven, A. C., 2.5
Stevens, E. D., 1.2
Stewart, A. T., 4.1
Stewart, R. F., 1.2
Stokes, H. T., 1.5
Stone, A. J., 3.5
Stoops, J. K., 2.5
Storks, K. H., 4.5
Storoni, L. C., 2.3
Stout, G. H., 1.3, 2.3
Stragler, H., 4.4
Strahs, G., 2.3
Strandberg, B., 2.3
Strandberg, B. E., 1.3, 2.3, 2.4
Strässler, S., 4.2
Stratonovich, R. L., 1.2
Stroud, R. M., 2.4
Stroud, W. J., 4.5
Strzelecki, L., 4.4
Stuart, A., 1.2, 2.1
Stuart, D., 2.3
Stubbs, G., 4.5
Stubbs, G. J., 4.5
Sturkey, L., 5.2
Sturtevant, J. M., 4.5
Su, Z., 1.2
Suck, D., 2.3, 3.3
Sugihara, K., 2.5
Sun, W., 2.5
Sundaralingam, M., 3.3
Sundaram, K., 3.3
Sundberg, M., 2.5
Suresh, K. A., 4.4
Suryan, G., 1.3
Suski, T., 4.1
Sussman, J. L., 1.3, 2.4
Sutcliffe, D. C., 3.3
Sutherland, I. E., 3.3
Süto, A., 2.2
Suzuki, E., 4.5
Suzuki, H., 4.5
Suzuki, S., 2.5
Svergun, D. I., 2.5
Swaminathan, S., 2.3, 3.3
Swanson, S. M., 3.3
Swartzrauber, P. N., 1.3
Swoboda, M., 4.3
Symmons, M. F., 4.5
Szegö, G., 1.3
Szillard, L., 2.4

Tadokoro, H., 4.5
Taftø, J., 4.3
Taftø, T., 2.5
Taguchi, I., 2.2
Tajbakhsh, A. R., 4.4
Takagi, K., 2.5
Takagi, S., 2.5, 5.1, 5.3
Takahashi, H., 5.2
Takahashi, M., 2.5
Takahashi, T., 5.3
Takaki, Y., 4.2
Takano, T., 1.3
Takayoshi, H., 2.5
Takenaka, A., 3.3
Takeuchi, Y., 2.3, 2.4
Talapov, A. L., 4.4
Tama, F., 2.5
Tamarkin, J. D., 1.3
Tanaka, I., 4.2
Tanaka, K., 1.2
Tanaka, M., 2.5, 5.2
Tanaka, N., 2.3, 2.4, 4.3
Tanaka, S., 4.5
Tang, G., 2.5
Tanji, T., 2.5
Tanner, B. K., 5.1
Tao, X., 2.3
Tao, Y., 2.3
Tardieu, A., 4.4
Tarento, R. J., 4.4
Tasset, F., 5.3
Tasumi, M., 3.3
Tatarinova, L. I., 4.5
Tate, C., 2.2, 2.3
Taupin, D., 5.3
Tavares, P., 2.5
Taveau, J. C., 2.5
Taylor, C. A., 1.3, 1.4, 4.2
Taylor, D. J., 2.2
Taylor, G. H., 4.4
Taylor, R., 3.1, 3.3
Taylor, W. J., 2.3
Tchoubar, D., 4.2
Teeter, M. M., 2.3, 2.4
Teller, A. H., 4.2
Teller, E., 4.2, 4.5
Temperton, C., 1.3
Templeton, D. H., 2.4
Templeton, L. K., 2.4
Ten Eyck, L. F., 1.3, 2.3
Teplyakov, A., 2.3
Terauchi, M., 2.5

Terwilliger, T. C., 2.2, 2.3, 2.4
Teworte, R., 5.1
Thierry, J. C., 2.3
Thoen, J., 4.4
Thomas, D. J., 3.3
Thomas, K. M., 4.4
Thomsen, K., 3.3
Thon, F., 2.5
Thorpe, M. F., 4.2
Thouless, D. G., 4.4
Thuman, P., 2.2
Tibbals, J. E., 4.2
Tikhonov, V. I., 1.2
Tildesley, D. J., 3.5
Timmer, J., 2.5
Tinh, N. H., 4.4
Tinnappel, A., 2.5
Tirion, M., 4.5
Titchmarsh, E. C., 1.3
Tivol, W. F., 2.5
Tivol, W. T., 4.5
Tjian, R., 3.3
Toby, B. H., 3.3
Toeplitz, O., 1.3
Tolimieri, R., 1.3
Tollin, P., 2.3
Tolstov, G. P., 1.3
Tomimitsu, H., 5.3
Toner, J., 4.4
Tong, L., 2.3
Toniolo, L., 4.2
Tonomura, A., 2.5
Torrisi, A., 3.5
Tosoni, L., 2.5
Toukmaji, A., 3.5
Toupin, R., 2.2
Tournaire, M., 2.5
Tournarie, M., 5.2
Tramontano, A., 3.3
Traub, W., 2.2
Trèves, F., 1.3
Trickey, S. B., 3.5
Trommer, W. E., 2.3
Tronrud, D. E., 1.3, 2.3
Trueblood, K. N., 1.1, 1.2, 1.3, 2.3
Trus, B. L., 2.5
Truter, M. R., 1.3
Tsai, A. P., 2.5
Tsao, J., 2.3
Tsernoglou, D., 3.3
Tsipursky, S. I., 2.5
Tsirelson, V. G., 2.5
Tsoucaris, G., 2.2, 2.4
Tsuda, K., 2.5
Tsuji, M., 2.5, 4.5
Tsujimoto, S., 2.5
Tsukihara, T., 2.3, 3.3
Tsuprun, V. L., 2.5
Tsuruta, H., 2.5
Tukey, J. W., 1.3
Tulinsky, A., 2.4
Turberfield, K. C., 4.2
Turkenburg, M. G., 2.3
Turner, J., 3.3
Turner, J. N., 2.5
Turner, P. S., 2.5, 4.5
Typke, D., 2.5

Uchida, Y., 4.3
Ueki, T., 2.4
Ueno, K., 2.5
Uhrich, M. L., 1.3
Ungaretti, L., 2.4
Unge, T., 2.3
Unser, M., 2.5
Unwin, P. N. T., 2.5
Uragami, T., 5.1
Urzhumtsev, A., 2.3
Usha, R., 2.3
Ushigami, Y., 5.3
Uson, I., 2.3
Utemisov, K., 5.3
Uyeda, N., 2.5
Uyeda, R., 2.5, 4.3

Vaara, I., 2.3
Vacher, R., 4.1
Vagin, A. A., 2.3, 2.5
Vainshtein, B. K., 2.5, 4.2, 4.5
Van Dael, W., 4.4
Van der Pol, B., 1.3
Van der Putten, N., 2.2
Van Hove, L., 4.1, 4.3
Van Loan, C. F., 2.5
Van Tendeloo, G., 4.3
Vand, V., 1.3, 2.5, 4.5
Varady, W. A., 4.4
Varghese, J. N., 1.3, 2.2
Varn, D. P., 4.2
Varnum, J. C., 2.3
Vartanyants, I. A., 5.1
Vaucher, C., 4.4
Vaughan, P. A., 2.2
Vedani, A., 3.3
Velazquez-Muriel, J., 2.5
Venkataraman, G., 4.1
Venkatesan, K., 2.4
Venuti, E., 3.5
Vereijken, J. M., 2.3
Vermin, W. J., 2.2
Vernoslova, E., 2.3
Verschoor, A., 2.5
Verwer, P., 3.5
Vibert, P. J., 4.5
Vicković, I., 2.2
Vijayan, M., 2.2, 2.3, 2.4
Vilkov, L. V., 2.5
Villain, J., 4.4
Vincent, R., 2.5
Vine, W. J., 2.5
Viterbo, D., 2.2
Vogel, R. H., 2.5
Volbeda, A., 2.3
Volkmann, N., 2.5
Volkov, A., 3.5
Von der Lage, F. C., 1.2
Vonderviszt, F., 4.5
Vonrhein, C., 2.3
Voronova, A. A., 2.5
Vos, A., 2.4
Vrána, M., 5.3
Vriend, G., 2.3
Vries, T. A. de, 2.3
Vrublevskaya, Z. V., 2.5
Vulis, M., 1.3

Wagenknecht, T., 2.5
Wagner, E. H., 5.1
Waho, T., 4.3
Waite, J., 3.5
Wakabayashi, K., 4.5
Walian, P. J., 2.5
Walker, C. B., 4.1
Waller, I., 1.2
Walsh, G. R., 3.3
Walz, J., 2.5
Wang, B. C., 1.3, 2.3, 2.4
Wang, D. N., 2.5
Wang, H., 4.5
Wang, J., 4.4, 5.1
Wang, R., 4.2
Wang, X. J., 4.4
Wang, Z. L., 4.3
Ward, J. C., 1.3
Ward, K. B., 2.3
Wark, J. S., 5.1
Warme, P. K., 3.3
Warren, B., 1.3
Warren, B. E., 1.3, 4.2, 4.4
Warren, G. L., 2.3
Warren, S., 4.5
Warshel, A., 3.3
Waser, J., 1.3, 1.4, 3.1, 3.2
Watanabe, D., 2.5, 4.2, 4.3
Watanabe, E., 2.5
Watenpaugh, K. D., 1.4, 2.4
Watkin, D. J., 3.3
Watson, D. G., 3.3
Watson, G. L., 1.3
Watson, G. N., 1.3
Watson, H. C., 2.3
Watson, K. J., 1.2
Watson, W. T., 2.5
Wawak, R. J., 3.5
Waxham, M. N., 2.5
Webb, H., 2.5
Weber, H. J., 3.5
Weber, H. P., 3.5
Weber, T., 4.2
Weber, Th., 4.2
Weckert, E., 5.1
Wedemeyer, W. J., 3.5
Weeks, C. M., 2.2, 2.3, 4.5
Weidner, E., 4.2
Weikenmeier, A., 2.5
Weintraub, H. J. R., 3.3
Weinzierl, J. E., 2.2, 2.3, 2.4
Weiss, A. H., 4.4
Weiss, G. H., 1.3, 2.1
Weiss, R., 2.3
Weiss, R. J., 1.2
Weissberg, A. M., 2.2
Welberry, T. R., 4.2, 4.5
Welch, A., 2.5
Welch, P. D., 1.3, 2.5
Wells, M., 1.3, 1.4
Welsh, L. C., 4.5
Wenk, H.-R., 2.5
Wentowska, K., 4.4
Werner, S. A., 4.2, 5.3
Wesolowski, T., 3.3
West, J., 1.3
Westbrook, J. D., 1.4
Westhof, E., 2.5
Weyl, H., 1.3
Weymouth, J. W., 4.1
Whelan, M., 4.3
Whelan, M. J., 2.5, 4.3, 5.2
White, C., 3.5
White, H., 2.5
White, J. G., 2.3
White, P., 2.2
Whitfield, H. J., 5.2
Whittaker, E. J. W., 1.3
Whittaker, E. T., 1.3, 3.2
Widder, D. V., 3.4
Widom, H., 1.3
Wiener, N., 1.3
Wigner, E. P., 1.5
Wiley, D. C., 2.3
Wilfing, A., 5.3
Wilke, S., 3.5
Wilke, W., 4.2
Wilkins, M. H. F., 4.5
Wilkins, S. W., 1.3, 2.2
Williams, D. E., 3.4, 3.5, 4.1
Williams, G. J. B., 3.3
Williams, R. M., 4.4
Williams, T. V., 3.3
Willingmann, P., 2.3
Willis, B. T. M., 1.2, 4.1, 4.6
Willoughby, T. V., 3.3
Wilson, A. J. C., 1.3, 2.1, 2.2, 2.3, 2.4, 4.2, 4.5
Wilson, E. B., 1.1
Wilson, I. A., 2.3
Wilson, K. S., 2.1, 2.2, 2.3
Wilson, S. A., 5.3
Windsor, C. G., 4.2
Wingfield, P. T., 2.5
Winkler, F. K., 1.3, 2.3
Winkor, M. J., 4.4
Winograd, S., 1.3
Winter, W. T., 4.5
Wintgen, G., 1.5
Wintner, A., 1.3
Wipke, W. T., 3.3
Withers, R. L., 2.5, 4.2
Witt, C., 2.5
Wittmann, J. C., 2.5, 4.5
Wolf, E., 5.1
Wolf, J. A., 1.3
Wolff, P. M. de, 2.2, 2.5, 4.2, 4.6
Wonacott, A. J., 2.4, 4.5
Wondratschek, H., 1.3, 1.4, 1.5, 2.2
Wong, H. C., 2.5
Wong, S. F., 4.2
Woods, R. E., 2.5
Woodward, I., 2.3, 2.4
Woolfson, M. M., 1.3, 2.1, 2.2, 2.3, 2.5
Wooster, W. A., 4.2
Wostrack, A., 4.2
Wright, D. C., 4.6
Wright, M. H., 3.3
Wrighton, P. G., 4.4
Wrinch, D. M., 2.3
Wu, H., 2.3
Wu, M. N., 2.5
Wu, T. B., 4.2
Wu, X.-J., 2.5
Wu, Y. K., 2.5
Wübbeling, F., 2.5
Wuensch, B. J., 4.2
Wunderlich, B., 4.5
Wunderlich, J. A., 2.3
Wyckoff, H. W., 1.3, 2.3, 4.5
Wynn, A., 3.3
Wynne, S. A., 2.5

Xia, D., 2.3
Xiang, S.-B., 2.5
Xiaodong, Z., 1.4
Xu, H., 2.2
Xu, P. R., 4.3
Xu, Y., 2.3

Yagi, N., 4.5
Yamamoto, A., 2.5, 3.3, 4.6
Yamamoto, N., 2.5
Yamashita, I., 4.5
Yamazaki, H., 5.1
Yang, C., 2.3, 2.5
Yang, W., 3.5
Yang, Y. W., 1.2
Yao, J.-X., 2.2, 2.5
Yeates, T. O., 2.3
Yelon, W. B., 5.3
Yessik, M., 4.2
Yin, Z. H., 2.5
Yin, Z. Y., 2.5
Yip, S., 4.1
Yonath, A., 2.2
York, D. M., 3.5
Yoshioka, H., 4.3
Yosida, K., 1.3
Youla, D. C., 2.5
Young, A. P., 4.4
Young, C. Y., 4.4
Young, R. A., 4.2
Yu, L. J., 4.4
Yuan, B.-L., 4.5
Yuen, C. K., 2.5

Zachariasen, W. H., 1.3, 1.4, 2.4, 5.1, 5.3
Zak, J., 1.5
Zalkin, A., 2.4
Zaluzee, N. J., 2.5
Zambianchi, P., 5.1
Zarka, A., 5.3
Zaschke, H., 4.4
Zassenhaus, H., 1.3
Zechmeister, K., 1.3, 2.5
Zegenhagen, J., 5.1
Zegers, I., 5.1
Zeilinger, A., 5.3
Zeitler, E., 2.5
Zelenka, J., 5.3
Zelepukhin, M. V., 5.3
Zellner, J., 5.1
Zelwer, C., 2.2
Zemanian, A. H., 1.3
Zemlin, F., 2.5
Zenetti, R., 4.2
Zeng, G. L., 2.5
Zernike, F., 4.2
Zeyen, C., 4.2, 5.3

AUTHOR INDEX

Zhang, K. Y. J., 2.3
Zhang, W. P., 2.5
Zhao, Z. X., 2.5
Zheng, C.-D., 2.2, 2.5
Zheng, Y. L., 2.5
Zhong, Z.-Y., 2.5
Zhou, Z. H., 2.5
Zhu, J., 2.5
Zhu, Y., 2.5
Zhukhlistov, A. P., 2.5
Zicovich-Wilson, C. M., 3.5
Ziman, J. M., 1.1
Zinn-Justin, J., 4.4
Zlotnick, A., 2.5
Zobetz, E., 4.6
Zolotoyabko, E., 5.3
Zou, J.-Y., 2.5, 3.3
Zucker, I. J., 3.4
Zucker, U. H., 1.2
Zugenmaier, P., 4.5
Zuo, J. M., 2.5
Zvyagin, B. B., 2.5
Zwick, M., 1.3, 2.3, 2.4
Zygmund, A., 1.3

Subject index

A posteriori probability, 512
A priori probability, 504
Ab initio phase determination, 273
Abbe theory, 303
Abel summation procedure, 46
Abelian groups, 42, 79
Aberrations, 303
Absolute configuration, 282, 285
Absolutely integrable functions, 27
Absorbing crystals, 629, 637–638, 640
Absorption coefficient, 634, 639
  effective, 628
  linear, 628
  phenomenological, 303
Absorption edge, 284
Absorption function, 542
Absorption in electron diffraction, 362
Accelerated convergence, 449
  formula *via* Patterson function, 454
Acceptance domain, 596
Acentric reflections, 74
Acoustic modes, 486
Action, 68
Additive reindexing, 61
Adiabatic approximation, 484
Adjusted coefficients, 180
Affine change of coordinates, 35
Affine change of variables, 40
Affine space-group type, 177
Affine transformation, 120
Agarwal's FFT implementation of the Fourier method, 98
Alfalfa mosaic virus, 259
Algebra of functions, 75
Algebraic integers, 78, 82
Algebraic number theory, 82
Algebraic reconstruction technique (ART), 370
Aliasing, 47, 49, 93
Alignment of electron-microscopy images, 379
Allowed origins, 215
'Almost everywhere', 26
Analytical methods of probability theory, 102
Angle between two vectors, 404
Angles
  Eulerian, 262, 420
  spherical, 262
Anisotropic displacement parameters, 443
Anisotropic displacement tensors, 6
Anisotropic fluid, 547
Anisotropic Gaussian atoms, 63
Anisotropic temperature factors, 73
Anisotropic weights, 413
Annular dark-field detector, 305
Anomalous absorption, 633, 659
Anomalous difference, 288, 291
Anomalous dispersion (scattering), 255–256, 282, 284
  integration with direct methods, 237
  Patterson function, 257
Anomalous scatterers, 64, 74, 256, 284, 286
Anomalous transmission effect, 633
Antiferromagnetic domains, 661
Anti-nodes of standing waves, 633
Antisymmetric tensor, 6
Aperiodic crystals, 590
  disorder diffuse scattering from, 526
  ideal, 590
Aperiodic structure, 590
Apparent noncrystallographic symmetry, 267
Approximate helix symmetry, 570
Approximations
  adiabatic, 484
  Bethe, second, 302
  Born, first-order, 10, 300
  Born, second-order, 11
  Born–Oppenheimer, 17
  Edgeworth, 21
  forward-scattering, 301
  harmonic, 484
  kinematical, 62, 300–301, 359, 394, 583, 658
  phase-grating, 652
  phase-object, 301, 542
  projected charge-density, 305
  projection, 648
  saddlepoint, 102–103
  seven-beam, 652
  small-angle-scattering, 300
  three-beam, 652
  two-beam, 302, 649
  two-beam dynamical, 394
  weak-phase-object, 304
Area detector, 294
Argand diagram, 283
Arithmetic classes, 71
  of representations, 71
Arithmetic crystal class, 177
Arms of star, 178
ART (algebraic reconstruction technique), 370
Artificial temperature factor, 94, 100
Aspherical-atom form factor, 14
Aspherical multipole refinement, 459
Associated actions in function spaces, 69
Associativity properties of convolution, 100
Assumption of independence, 209
Assumption of uniformity, 203, 209
Astigmatism, 378
Asymmetric carbon atom, 285
Asymmetric images, 365
Asymmetric unit, 69, 72, 180
  noncrystallographic, 253
Asymmetry ratio, 631
Asymptotic crystal shape, 470
Asymptotic distribution of eigenvalues of Toeplitz forms, 45, 68
Asymptotic expansions
  and limit theorems, 103
  of Gram–Charlier and Edgeworth, 105
Atom-centred spherical harmonic expansion, 12
Atomic characteristic functions, 211
Atomic electron densities, 75
Atomic error matrix, 416
Atomic force-constant matrix, 485
Atomic form factor, 10
  X-ray, 293
Atomic scattering factor, 10, 284
  spherical, 10
Atomic scattering length, 11
Atomic surface, 590, 596
Atomic temperature factor, 17
ATOMS, 443–445
Autocorrelation, 65
Autocorrelation function, 388
Automated Patterson-map search, 389
Automorphism, 69, 71
Auxiliary basis set expansion, 479
Auxiliary basis set fitting, 459
Average difference cluster method, 518
Average intensity
  of general reflections, 195
  of zones and rows, 196
Average multiples for point groups, 197
Averaged electron density, 274
Axial disorder, 517

B-splines, 476
  approximation to trigonometric functions, 477
  derivatives of, 477
  Fourier transforms of, 477
  recursion for, 477
  tensor product, 480
B3LYP basis set, 479
Back-shift correction, 96
Back surface, 639
Background diffraction, accurate subtraction of, 575
Backprojection, 367
  filtered, 371
Backward convolution theorem, 44, 75
Bacterial rhodopsin, 274
Balls&Sticks, 443–444
BALSAC, 443–445
Banach spaces, 28
Band-limited function, 49
Base-centred lattices, 90
Bases
  Cartesian, 7
  contravariant, 5–6
  covariant, 5–6
  direct and reciprocal, relationships between, 3
  mutually reciprocal, 2–3
  primitive, 177
  reference, choice of, 7
Basic crystallographic computations, 91
Basic domain, 180
Basic structure, 591
Basis vectors, contravariant, 405
Bayesian statistical approach to the phase problem, 106
Beevers–Lipson factorization, 58, 76–77
Beevers–Lipson strips, 76, 93
Bessel's inequality, 47
Best Fourier, 91, 290
Best phase, 290
Best plane, 410
Beta distribution
  first kind, 201
  second kind, 201
Bethe approximation, second, 302
Biaxial nematic order, 549
Bieberbach theorem, 68
Bijvoet differences, 257
Bijvoet equivalents, 286–287
Bijvoet pair, 285–286
Bilder, 440
Binary systems, distortions in, 522
Binding energy, 284
Bloch-wave formulation, 650
Bloch waves, 300, 628
Bloch's theorem, 9, 486
  alternative form of, 9
Blow and Crick formulation, 290
Body-centred lattices, 90
Body-diagonal axes, 128
Bond angles, 437
Bond orientational order, 547
Booth's differential Fourier syntheses, 96
Booth's method of steepest descents, 96
Bootstrap technique, 388
Borie–Sparks method, 523
Born approximation
  first-order, 10, 300
  second-order, 11
Born–Oppenheimer approximation, 17
Born series, 300, 651
  expansion, 543
Born–von Karman boundary conditions, 177
Born–von Kármán theory, 484
Borrmann effect, 303, 633
Borrmann triangle, 641
Boundary conditions, 628, 639
  at exit surface, 635
  periodic, 479
Bounded projections, 67, 92
Bounded subset, 25
Boy's function, 471
BP (bright-field pattern), 310
Bragg case, 631
Bragg's law, departure of incident wave from, 630
Bragg–Lipson charts, 93
Branch, 485
Bravais lattices
  centred, 121
  direct and reciprocal, 121
Bright-field image intensity, 362
Bright-field pattern (BP), 310
Brillouin zone, 9, 485
  first, 178
Bulk plasmon excitation, 300
Burg entropy, 68
Burnside's theorem, 71
Butterfly loop, 53

Calculus
  of asymmetric units, 79
  operational, 28
Cameron, 443–444
CaRIne, 443–444
Carpet of cross-vectors, 260
Cartesian basis, 7
Cartesian coordinate system, 404
Cartesian coordinates, 418, 445
Cartesian frames of reference, 5
Cartesian product, 25, 42
Cartesian system, transformation to, 3
Cauchy kernel, 46
Cauchy–Schwarz inequality, 27, 47
Cauchy sequence, 26
Cauchy's theorem, 104
CBED (convergent-beam electron diffraction), 307
  coherent, 323, 333
CCP14 (Collaborative Computational Project Number 14), 443
Cell constants, 576
Central-limit theorem, 103, 198
  Lindeberg–Lévy version, 204
Central section theorem, 367
Centre-of-mass translational displacements, 525
Centre of symmetry, false, 115
Centred Bravais lattice, 121
Centred lattices, 73
Centric reflections, 73
Centring
  effect of, 196
  translations, 121
  type, 121
Centrosymmetric projections, 251–252
Centrosymmetry, status of, 123
Cesàro sum, 46
Chain rule, 99
Chain trace, 441
Chains, flexible, 436
Change-of-basis matrix, 123
Change of crystal axes, 120
Channelling pattern, 544
Characteristic functions, 102, 197, 208
  atomic, 211
Charge densities, Coulomb energy of, 474
Charge distributions, Gaussian, 461, 471–472
CHARMM, 442
CHEMGRAF, 439
Chemical correctness of polypeptide fold, 273
Chem-X, 442
Chinese remainder theorem (CRT), 54, 61, 82
  for polynomials, 57, 84
  reconstruction, 54
  reconstruction formula, 57
Chirality, 285
Choice of reference bases, 7
Cholesteryl iodide, 249
α-Chymotrypsin, 271
Circular harmonic expansions, 100
Classical Thomson scattering, 10
Classification of crystallographic groups, 70
Classification of electron-microscopy images, 381
Clebsch–Gordan coefficients, 17
Closed point group, 258
Closed subset, 25
Cluster model, 545
Clustering, 518
Clustering algorithms, 380
Clusters, 500
Cochran's Fourier method, 96
Cocycle, 85
Coherence, 434
Coherence length, 568
Coherent convergent-beam electron diffraction, 323, 333
Coherent scattering, 488
Collaborative Computational Project Number 14 (CCP14), 443
Column part, 177
Common line, 382
Communication, statistical theory of, 104
Commutative ring, 54
Compact Gaussians, 472
Compact subset, 25
Compact support, 25, 36, 44
  distributions with, 30, 40, 43–44, 47
Complement of the incomplete gamma function, 450–451
Complete normed space, 28
Complete vector spaces, 26
Completely reducible matrix group, 176
Complex antisymmetric transforms, 87
Complex scattering factor, 255
Complex symmetric transforms, 87
Components of vector products, 405
Components of vectors, 5
Composite lattice, 449, 452
Composite structure, 593
Compound nucleus, 11
Compound transformations, 429
Compton scattering, 489
Computational and algebraic aspects of mutually reciprocal bases, 4
Computational cost of Ewald direct sum, linear scaling of, 474
Computer-adapted space-group symbols, 117, 122, 127
Computer-algebraic languages, 122
Computer architecture, 52, 61
Computer simulations of diffuse scattering, 528
Condensed ring systems, 437
Conditional convergence, 460
Conditional pair probability, 504
Conformational variability, 388
Conforming/nonconforming disorder, 518
Conformons, 431
Conjugacy classes of subgroups, 69
Conjugate and parity-related symmetry, 85
Conjugate distribution, 104–105
Conjugate families of distributions, 106
Conjugate gradient method, 436
Conjugate symmetry, 35, 40
Conjugation, 69
Connectivity
  drawing, 435
  implied, 435
  logical, 435
  structural, 435
Connectivity tables, 435
Connectivity tree, 439
Consistency condition, 39
Constant **Q** mode, 490
Constraints, 436, 440–441
  on interpretation of Patterson functions, 258
Continuous diffraction on layer lines, 575
Continuum dielectric medium, 467
Contragredient, 40
  of a matrix, 35
Contrast transfer function, 370, 376
Contravariant bases, 5–6
Contravariant basis vectors, 405
Contravariant components, 5
Conventional coefficients, 180
Convergence
  accelerated, 449
  accelerated, formula *via* Patterson function, 454
  conditional, 460
  of distributions, 30
  of Fourier series, 45
Convergence-accelerated direct sum, 452
Convergence method, 231
Convergent-beam electron diffraction (CBED), 307
  coherent, 323, 333
Conversion of translations to phase shifts, 35
Convolution, 64, 244
  associativity properties of, 100
  cyclic, 51, 55
  of distributions, 34
  of Fourier series, 44
  of probability densities, 102
  of two distributions, 34
Convolution property, 35, 51
Convolution techniques, 392
Convolution theorem, 37, 42, 44, 46, 67, 102, 106
  backward version, 44, 75
  forward version, 44, 64–65, 75
Convolution theorems with crystallographic symmetry, 75
Cooley–Tukey algorithm, 52, 61, 77
  vector-radix version, 58
Cooley–Tukey factorization, multidimensional, 58–59, 79
Coordinate systems, 177
  Cartesian, 404
  natural, 404
Coordinates
  affine change of, 35
  Cartesian, 418, 445
  crystallographic, 418
  fractional, 42, 63, 262
  homogeneous, 418, 421
  nonstandard, 42
  positional, 437
  screen, 426, 428
  spherical, 467
  spherical polar, 263
  standard, 42, 63, 71–72
  transformation of, 5, 7, 33
  world, 426
Copolymers, random, 571
Core of discrete Fourier transform matrix, 83
Correction-factor approach, 203, 212
Correlated lattice disorder, 573
Correlation, 64
Correlation functions, 75, 100, 252, 489, 505, 541
  Ornstein–Zernike, 514
  short-range-order, 518
Correlation length, 550
  pretransitional lengthening of, 551
Correlations
  intermolecular, 525
  librational–librational, 525
  vibrational–librational, 525
Coset averaging, 48–49
Coset decomposition, 48, 58
Coset reversal, 59
Cosets, 48, 71
  left, 68
  right, 69
Cosine strips, 76
Coulomb energy, 449
  of charge densities, 474
Coulomb interactions, 458
  between Gaussians, 459
  damped, 471
  direct, 472
Coulombic lattice energy, 449, 453
Covariance, 407
  interatomic, 411
Covariances, 411
Covariant bases, 5–6
Covariant components, 5
Cowpea mosaic virus, 259
Critical angle, 554
Critical scattering, 551
Cross correlation, 76, 388
Cross-correlation function, 365
Cross-Patterson vectors, 260
Cross-rotation function, 100
Cross-vectors, 251
  carpet of, 260
CRT (Chinese remainder theorem), 54, 61, 82
  for polynomials, 57, 84
  reconstruction, 54
  reconstruction formula, 57
Cruickshank's modified Fourier method, 97
Cryo-electron microscopy (cryo-EM), 375
Cryscon, 443
Crystal axes, change of, 120
Crystal class, arithmetic, 177
Crystal defects in thin films, 542
Crystal periodicity, 62
Crystal-structure imaging, 306
Crystal-structure prediction, 458
Crystal structures
  display of, 443
  incommensurate, 445
  magnetic, 445
  polyhedral display of, 443
  root-mean-square differences between, 445
Crystal Studio, 443–444
Crystal symmetry, 68

Crystal systems, 71, 123
Crystal-B phase, 558
Crystal-E phase, 560
Crystal-G phase, 560
Crystal-H phase, 560
Crystal-J phase, 560
Crystal-K phase, 560
Crystalline approximant, 590
Crystallographic applications of Fourier transforms, 62
Crystallographic coordinates, 418
Crystallographic discrete Fourier transform, 77
  algorithms, 76
Crystallographic extension of the Rader/Winograd factorization, 82
Crystallographic Fourier transform theory, 62
Crystallographic group action, 79
  in real space, 71
  in reciprocal space, 72
Crystallographic groups, 68
  classification of, 70
Crystallographic statistics, 204
Crystallographic symmetry, 258
Crystallographica, 443–444
CrystalMaker, 443–445
CrystMol, 443–445
Cubic space groups, 90, 118
Cumulant expansion, 21
Cumulant-generating functions, 103, 198
Cumulative distribution functions, 201
Cuprous chloride azomethane complex, 247
Cyclic convolution, 51, 55
Cyclic (even) permutation of coordinates, 118, 122
Cyclic groups, 71
Cyclic symmetry, 83
Cyclotomic polynomials, 57
Cylindrically averaged diffraction patterns, 572
Cylindrically averaged Patterson function, 577

Damped Coulomb interaction, 471
Dark-field pattern (DP), 309
Data flow, 61
Data handling, Hall symbols in, 123
Data space, 426–427, 429–430
de la Vallée Poussin kernel, 46
Debye model, 486
Debye theory, 484
Debye–Waller factor, 551, 634
Decagonal phase, 607
Decagonal point groups, 354
Decagonal quasicrystals, 354
Decimation, 25, 49, 53
  and subdivision of period lattices, duality between, 48
  in frequency, 54, 59, 86
  in time, 53, 85
  period, 48
Decimation matrix, 58–59, 78
Decomposition, 78
  coset, 48, 58
  orbit, 69, 72, 74–75
Deconvolution of a Patterson function, 249
Defects, 299, 518
Defocus, 366
  optimal, 361
  Scherzer, 305
  Scherzer, conditions, 362
Deformed crystal, 632
Delta functions, 25
  Dirac, 28, 450
  periodic, 211
  three-dimensional Dirac, 464
  transforms of, 40
Density modification, 91, 392
Density modulation, 601
  harmonic, 601
  symmetric rectangular, 601
Density of nuclear matter, 2
Deoxyhaemoglobin, 273
Depth cueing, 429, 439
Derivatives
  for model refinement, 95
  for variational phasing techniques, 94
  of B-splines, 477
Detectors
  annular dark-field, 305
  area, 294
Determinantal formulae, 227
Determinantal inequalities, 67
Deviation parameter, 631, 633
Diamond, 443–445
Diamond's real-space refinement method, 99
Dielectric response, 468
Dielectric susceptibility, 627, 642
  Fourier expansion of, 627
Difference Fourier analysis, 576
Difference Fourier synthesis, 288, 579
Difference Fourier technique, 288
Difference Patterson functions, 253–254
  isomorphous, 253
Differential syntheses, 36, 67, 97
Differentiation, 25, 36
  and multiplication by a monomial, 40
  of distributions, 31
  under the duality bracket, 31
Differentiation identities, 51
Differentiation property, 106
Diffraction, dynamical, 393
Diffraction beams, intensities of, 358
Diffraction by helical structures, 100, 568
Diffraction conditions, 63
Diffraction groups, 308–311, 313
Diffraction imaging techniques, 661
Diffraction patterns, cylindrically averaged, 572
Diffraction relations, 2
Diffraction vector, 2
Diffractometers, optical, 364
Diffuse Gaussians, 472
Diffuse scattering, 540
  computer simulations of, 528
  elastic, 492
  from aperiodic crystals, 526
  from polycrystalline materials, 534
  from quasicrystals, 527
  modelling of, 528
  of X-rays, 492
Digit reversal, 53, 59, 62
Digital electronic computation of Fourier series, 76
Dihedral symmetry, 84
Dimension of a representation, 176
Dipalmitoylphosphatidylcholine, 562
Dipole moment, 454
  of unit cell, 461
Dirac delta function, 28, 450
  three-dimensional, 464
Direct Bravais lattice, 121
Direct Coulomb interaction, 472
Direct Fourier inversion, 370
Direct inspection of structure-factor equation, 116
Direct lattice, 2, 5, 42, 121, 177
Direct-lattice sum, 453
Direct methods, 102, 117, 215, 288
  in macromolecular crystallography, 235
  integration with anomalous-dispersion techniques, 237
  integration with isomorphous replacement techniques, 236
Direct-methods packages, 234
Direct metric, 4
Direct phase determination, 36
  in electron crystallography, 388
Direct space, 176
Direct-space crystal lattice, 450
Direct-space sum, 449
Direct-space transformations, 120
Direct sum potential, 462
Direct unit-cell parameters, 4
Direction cosines of plane normal, 414
Dirichlet kernel, 46
  spherical, 64, 91
Discotic phases, 561
Discrete Fourier transform matrix, core of, 83
Discrete Fourier transformation, 47
Discrete Fourier transforms, 24, 77
  algorithms, 76
  matrix representation of, 51
  numerical computation of, 52
  properties of, 51
Discretization, 368
Dislocations, 555
Disorder, 540
  axial, 517
  conforming/nonconforming, 518
  from turns, twists and torsions of chains, 517
  lattice, 572
  lattice, correlated, 573
  longitudinal, 516
  orientational, 524
  substitutional, 545, 572
  two-dimensional, 514
Disordered fibres, 568
Dispersion corrections, 284, 286, 627
Dispersion effects, 284
Dispersion energy, 449
Dispersion equations, 300
Dispersion interactions, 459
Dispersion surface, 629, 650, 657
Displacive modulation, 600
  harmonic, 601
Display space, 426–427, 429–431, 434
Displaying crystal structures, 443
Distance function, 27
Distribution function, 509
  cumulative, 201
Distributions
  associated with locally integrable functions, 30
  beta, first kind, 201
  beta, second kind, 201
  conjugate, 104–105
  conjugate families of, 106
  convergence of, 30
  convolution of, 34
  definition of, 30
  differentiation of, 31
  division of, 32
  electron-magnetization, 11
  equal, 45
  Fourier transforms of, 39
  gamma, 201
  Gaussian, 290
  hypersymmetric, 201
  ideal acentric, 200
  ideal centric, 200
  integration of, 32
  lattice, 43–44, 47–48
  maximum-entropy, 36, 106
  moments of, 102
  motif, 63
  multiplication of, 32
  non-ideal, 203, 207
  of finite order, 30
  of random atoms, 104
  of sums, averages and ratios, 202
  operations on, 31
  periodic, 42, 44, 62
  probability density, 197
  probability density, ideal, 200
  support of, 30
  T on Ω, 30
  tempered, 36, 39, 41, 47, 72
  tensor products of, 33
  theory of, 25, 28
  with compact support, 30, 40, 43–44, 47
Divided differences, 651
Division of distributions, 32
Division problem, 32
Docking, 387, 441
Domain
  basic, 180
  minimal, 179
  of influence, 178
  representation, 179
Domain structure, 441
Double-phased synthesis, 283
Double-sorting technique, 273
DP (dark-field pattern), 309
Drawing connectivity, 435
DrawXTL, 443–445
Dual, topological, 30, 39–40
Dual relationships, 2

Duality
between differentiation and multiplication by a monomial, 67
between periodization and sampling, 44
between sections and projections, 41
between subdivision and decimation of period lattices, 48
Duality bracket, 39
Duality product, 31
Dummy indices, 5
Dynamic parallax, 427, 441
Dynamical approximation, two-beam, 394
Dynamical diffraction, 393, 542
theory, 301, 626
two-beam, formulae, 303
Dynamical extinction, 319, 323–324, 333, 335
Dynamical matrix, 486
Dynamical scattering effects, 388
Dynamical scattering factor, 543
Dynamical shape function, 651
Dynamical theory, 301, 626
fundamental equations, 628
of neutron diffraction, 654
plane-wave, 630
solution of, 633
Dynamics, 9
of three-dimensional crystals, 484

E maps, interpretation of, 232
Edgeworth approximation, 21
Edgeworth series, 103
EDSA (electron-diffraction structure analysis), 356
Effect of centring, 196
Effective absorption coefficient, 628
Effective potential-energy function, 581
Effects of symmetry on the Fourier image, 114
Eigenspace decomposition of L^2, 37
Eigenvalue, 649
Eigenvalue decomposition, 436
Eigenvalues and eigenvectors of orthogonal matrices, 425
Einstein model, 486
Elastic component of X-ray scattering, 10
Elastic constants, measurement of, 490
Elastic diffuse scattering, 492
Electromagnetic electron lenses, 299, 303
Electron band theory of solids, 629
Electron crystallography, 388
direct phase determination, 388
of polymers, 583
of proteins, 389
three-dimensional structure determination by, 391
Electron density, 2, 8, 15, 115, 290
averaged, 274
real-space averaging of, 259, 273
Electron-density calculations, 74
Electron-density maps, Fourier synthesis of, 91
Electron diffraction, 540
absorption in, 362
sign conventions, 301–302
Electron-diffraction data
for crystal-structure determination, 583
three-dimensional, 391, 585
Electron-diffraction patterns
geometric theory of, 359
polycrystal, 358
single-crystal, 356
texture, 357, 394
Electron-diffraction structure analysis (EDSA), 356
Electron distribution, atomic, radial dependence of, 12
Electron lenses, electromagnetic, 299, 303
Electron-magnetization distribution, 11
Electron micrographs
Fourier transform of, 393, 584
phase information from, 390
Electron-microscope image contrast, 542
Electron-microscope imaging, 541
Electron tomography, 377, 382
Electronic analogue computer X-RAC, 76
Electronic structure, 9
Electrons, interaction with matter, 299
Electrostatic energy, 449, 463
Electrostatic potential, 2, 461
Electrostatic properties of molecular surfaces, 438
Embedding method, n-dimensional, 591
Enantiomer, 286
Enantiomeric ambiguity, 247
Enantiomorph definition, 231
Enantiomorphic images, weak, 247
Enantiomorphic solutions, 246
Energy minimization, 437, 440, 442
Entire functions, 36
Entrance surface, 639
Entropy, 94
Envelope, 258
Envelope functions, 305
Epitaxic orientation techniques, 583
Equal-amplitude assumption, 580
Equal distribution, 45
Equivalent matrix groups, 176
Equivalent reflections, 285
Error matrix, atomic, 416
Error propagation, 411
Errors, 289
root-mean-square, 291
systematic, 408
Essential bounds, 45
Essentially bounded function, 27
Euclidean algorithm, 48, 56, 66
Euclidean norm, 25
Euclidean space, 25
Euler gamma function, 463
Euler spline, 477
Eulerian angles, 262, 420
Eulerian space, 263
Eulerian space groups, 263
rotation-function, 265
Eulerian symmetry elements, 264
Even (cyclic) permutation of coordinates, 118, 122
Ewald method, 458
Ewald potential, 466
Ewald result, 449
Ewald sum surface term, 467
Ewald wave, 628
Exchange between differentiation and multiplication by monomials, 102
Exchange between multiplication and convolution, 25
Excitation error, 647
Excitations
bulk plasmon, 300
inner-shell, 300
interband, 300
intraband, 300
Explicit-origin space-group notation, 127
Explicit space-group symbols, 123–124
Exploration of parameter space by molecular model building, 576
Exponential coefficient, 13
Exsolution, 505
Extended resolution, 274
Extended structures, 445
External fields, effect on neutron scattering, 659
External modes, 486
Extinction, 638
dynamical, 319, 323–324, 333, 335
Extinction distance, 632–633
Extinction factors, 658

F_{HLE}, 288
F_{HUE}, 288
Face-centred lattices, 90
Face-diagonal axes, 128
Factor group, 69, 71
Factorization, 78
False centre of symmetry, 115
Fast Fourier Poisson method, 475, 478
Fast Fourier transform (FFT), 77
three-dimensional (3DFFT), 475–476
Fast rotation function, 268
Feedback method, 253
Fejér kernel, 46, 464
spherical, 64
Fermi pseudo-potential, 654
Fermi surface, 545
FFT (fast Fourier transform), 77
three-dimensional (3DFFT), 475–476
Fibonacci chain, 595
Fibonacci sequence, 594
Fibre axis, 568
Fibre diffraction, 41, 576
R factor, 582
specimens for, 568
X-ray, 567
Fibres
axially periodic, transform of, 101
disordered, 568
macromolecular, 581
noncrystalline, 568, 570, 576
partially crystalline, 572
polycrystalline, 568, 570, 576
Field emission gun, 307
Figures of merit, 232, 290
Films
freely suspended, 554
smectic, 554
Filtered backprojection, 371
Filtered image, 364
Filtering
iterative low-pass, 575
rotational, 365
Finite crystal of point charges, 465
Finite field, 55
Finite space group, 177
Finite spherical crystal, 469
First Brillouin zone, 178
First-order Born approximation, 10, 300
First-order perturbation theory, 411
Flagpole, 183
Flexible chains, 436
Flexible rings, 436
Flight time, neutron, 660
Flipping ratio, 658
Focusing of neutron beams, 660
Force-constant matrix, atomic, 485
Form factor, 63
aspherical-atom, 14
atomic, 10
atomic, X-ray, 293
geometric, 605
Kikuchi-line, 543
FORTRAN, 122
FORTRAN interface, 122
FORTRAN interpreter, 122
Forward convolution theorem, 44, 64–65, 75
Forward scattering, 647
Forward-scattering approximation, 301
Four-dimensional vector, 424
Fourier analysis, 63
and filtration in reciprocal space, 364
Fourier approach, 212
Fourier–Bessel series, 209
Fourier–Bessel structure factors, 569
Fourier coefficients, 46, 283, 464
Fourier convolution theorem, 10
Fourier cotransform, 35
Fourier cotransformation, 41
Fourier expansion, 2
of dielectric susceptibility, 627
Fourier images, 114, 306
effects of symmetry on, 114
Fourier inversion, direct, 370
Fourier map, 283
Fourier method, 207
Agarwal's FFT implementation of, 98
Cochran's, 96
Cruickshank's modified, 97
Fourier representation, 293
Fourier series, 24
convergence of, 45
convolution of, 44
digital electronic computation of, 76
electron density and its summation, 63
partial sum of, 46
Fourier shell correlation (FSC), 384–385
Fourier space, 116
symmetry in, 120
Fourier summations, 116
space-group-specific, 116

Fourier synthesis, 63, 286, 290
  best, 290
  of electron-density maps, 91
Fourier-transform space, 450
Fourier transformation
  discrete, 47
  inverse, 36, 41
  mathematical theory of, 24
Fourier transforms, 24, 35, 450
  crystallographic applications, 62
  crystallographic, discrete, 77
  crystallographic, theory of, 62
  discrete, 24
  discrete, core of matrix, 83
  discrete, matrix representation of, 51
  discrete, numerical computation of, 52
  discrete, properties of, 51
  exchange of subdivision and decimation, 49
  in L^1, 35
  in L^2, 37
  in polar coordinates, 101
  in $\mathscr{S}$, 37
  inverse, 8
  kernels of, 35
  of a distribution, 39
  of a Gaussian, 463
  of B-splines, 477
  of density, 478
  of electron micrographs, 393, 584
  of periodic distributions, 42
  of tempered distributions, 39–40
  tables of, 39
  tensor product property of, 76
  various writings of, 39
FpStudio, 443–445
Fractal atomic surface, 597
Fractal sequence, 598
Fractional coordinates, 42, 63, 262
Fréchet space, 28
Freely suspended films, 554
Fresnel reflection law, 554
Friedel equivalent, 282, 285
Friedel pair, 285
Friedel's law, 64, 73, 75, 255, 299
Frobenius congruences, 71, 73
Frodo, 441
FSC (Fourier shell correlation), 384–385
Fubini's theorem, 27, 35, 39
Function spaces
  associated actions in, 69
  topology in, 27
Functional derivative, 99
Functions of polynomial growth, 41
Fundamental domain, 68–70, 72
Fundamental equations of dynamical theory, 628
Fundamental region, 178
Fundamental relationships, 3
Fused-ring systems, 442

G-invariant function, 70
Gamma distribution, 201
Gamma functions, 450
  Euler, 463
  incomplete, 450, 454
Gamma radiation, 293
Gauss' law, 461
Gaussian atomic densities, 39
Gaussian atoms, 72–73, 93
  anisotropic, 63
Gaussian charge distributions, 461, 471–472
Gaussian density, sampling of, 479
Gaussian distribution, 290
Gaussian function, 39
  standard, 38, 41
Gaussian plane, general, 413
Gaussian weights, 413
Gaussians, 99
  compact, 472
  Coulombic interactions between, 459
  diffuse, 472
  Fourier transform of, 463
  interacting spherical, lattice sums of, 471
  normalized, 472
General conditions for possible reflections, 115
General Gaussian plane, 413
General **k** vector, 178
General linear change of variable, 35
General multivariate Gaussians, 38
General reflections, average intensity of, 195
General superposition, 422
General topology, 27
General transformation, 423
General translation function, 269
Generalized multiplexing, 88
Generalized Patterson function, 494
Generalized Rader/Winograd algorithms, 90
Generalized structure-factor formalism, 22
Generalized support condition, 34
Geometric form factor, 605
Geometric redundancies, 66
Geometric structure factors, 116–117, 135
Geometric theory of electron-diffraction patterns, 359
Gibbs phenomenon, 46, 64
GKS (Graphical Kernel System), 419
GKS-3D (Graphical Kernel System for Three Dimensions), 419
Global crystallographic algorithms, 89
Global minimum, 437–438
Glyceraldehyde-3-phosphate dehydrogenase, 259, 273–274
Good algorithm, 54
Good factorization, multidimensional, 82
Goodness of fit, 416
Gram–Charlier series, 20, 103
Gram–Schmidt process, 425, 437
GRAMPS, 439
Graphical Kernel System (GKS), 419
Graphical Kernel System for Three Dimensions (GKS-3D), 419
Graphics, 418
  standards for, 419
Gravity, 660–661
Green's theorem, 32, 93
GRETEP, 443–445
Gridding method, 373
GRIP, 440
Group actions, 68, 77
  crystallographic, 79
  crystallographic, real space, 71
  crystallographic, reciprocal space, 72
Group characters, 89
Group cohomology, 80
Group extensions, 71
Group of units, 52
Group ring
  integral, 79
  module over, 24
Group–subgroup relationship, 118
Groups, 68
Guide, 441

Haemoglobin, 251–252, 282, 287
  horse, 252
Hall symbols, 123, 127, 130
  in data handling, 123
  in software, 123
Handedness, 333
Hankel transform, 101
Hardy's theorem, 39
Harker diagram, 257, 283, 290
Harker lines, 249
Harker peaks, 76
Harker planes, 249
  special, 249
Harker sections, 248–249
Harmonic approximation, 484
Harmonic density modulation, 601
Harmonic displacive modulation, 601
HDD (high-dispersion diffraction), 356
Heavy-atom-derivative data sets, scaling of, 255
Heavy-atom derivatives, 287
Heavy-atom distribution, 288
Heavy-atom location, 249
  three-dimensional methods, 252
Heavy-atom lower estimate, 257
Heavy-atom parameters, 289
Heavy-atom sites, 251
Heavy-atom substitution, 255
Heavy atoms, 286
HEED (high-energy electron diffraction), 356
Heisenberg's inequality, 39, 91
Helical structures, 518
  diffraction by, 100, 568
Helical symmetry, 101, 375, 569, 576
  approximate, 570
Helix repeat units, 569
Hermann–Mauguin space-group symbol, 119
Hermite coefficients, 476
Hermite function, 38, 103
  multivariate, 38, 99
Hermite Gaussians, 473
  normalized, 473
Hermite polynomials, three-dimensional, 21
Hermitian-antisymmetric transforms, 87
Hermitian form, 44
Hermitian symmetry, 64, 74, 85
Herringbone packing, 560
Hexagonal axes, 118
Hexagonal family, 118
Hexagonal space groups, 90, 119
Hexatic phase, 555–556
  in two dimensions, 555
  tilted, 556
Hexatic-B phase, 558
Hexokinase, 273
Hidden-line algorithms, 434
Hidden-surface algorithms, 434
High-dispersion diffraction (HDD), 356
High-energy electron diffraction (HEED), 356
High-resolution electron diffraction (HRED), 356
High-resolution electron microscopy (HREM), 360
High-voltage limit, 652
Higher angular momentum charge distributions, 473
Higher-order Laue zone (HOLZ) reflections, 308, 335
Highlighting, 434
Hilbert space, 27, 35, 47
Hologram, in-line, 307
Holohedral point group, 179
Holosymmetric space group, 180
HOLZ (higher-order Laue zone) reflections, 308, 335
Homogeneous coordinates, 418, 421
Homogeneous symmetric polynomial, 652
Homometric pair, 246
Homometric structures, 246
Homomorphism, 176
Horse haemoglobin, 252
HRED (high-resolution electron diffraction), 356
HREM (high-resolution electron microscopy), 360
Hybridization, 442
HYDRA, 441
Hydrogen bonding, 439, 441–442
Hydrophobic properties of molecular surfaces, 438
Hyperatoms, 592
Hypercrystal, 590, 592
Hypersymmetric distributions, 201
Hypothetical atoms, 94

Icosahedral phase, 613
Icosahedral point groups, 353
Icosahedral quasicrystals, 352
Ideal acentric distributions, 200
Ideal aperiodic crystal, 590
Ideal centric distributions, 200
Ideal crystal, 176, 590
Ideal probability density distributions, 200
Idempotents, 54
Image averaging in real space, 364
Image contrast, electron-microscope, 542
Image detection, 249
Image enhancement, 361, 363
Image intensity, bright-field, 362
Image of a function by a geometric operation, 26
Image processing in transmission electron microscopy, 361
Image reconstruction, 361
Image resolution, 306
Image restoration, 361–362
Images
  asymmetric, 365
  filtered, 364
  Fourier, 306

with point symmetry, 365
Immunoglobulin, 271
Implication theory, 248
Implicit function theorem, 33
Implied connectivity, 435
Improper rotation axes, 258
Improper rotations, 123, 128
In-line hologram, 307
Incident wave, departure from Bragg's law, 630
Incoherent inelastic scattering, 488
Incoherent scattering, 488
Incommensurability, 550
Incommensurate crystal structures, display of, 445
Incommensurate intergrowth structure, 593
Incommensurately modulated structures, 344, 591
Incomplete gamma function, 450, 454
complement of, 450–451
evaluation of, 454
Independence, assumption of, 209
Index, 68
Index of refraction, 628
Indicator functions, 32, 42, 47, 65, 91–92
Induction formula, 105
Inductive limit, 30
Inelastic component of X-ray scattering, 10
Inelastic scattering, 300, 540
neutron, 488
X-ray, 490
Inequalities among structure factors, 221
Inner-shell excitations, 300, 542
Insight, 439
Insight II, 442
Instrumental resolution, 307
Integral group ring, 79
Integral representation, 68
theory, 71
Integrals
Lebesgue, 26
Riemann, 26
Integrated intensity, 637–638, 640
Integration
by parts, 31
Lebesgue's theory of, 28
of anomalous-dispersion techniques with direct methods, 237
of distributions, 32
of isomorphous replacement techniques with direct methods, 236
Intensities of diffraction beams, 358
Intensities of plane waves
in reflection geometry, 638
in transmission geometry, 634
Intensities of reflected and refracted waves, 634
Intensity differences, 285
Intensity statistics, 105
Interaction between symmetry and decomposition, 78
Interaction between symmetry and factorization, 79
Interaction matrix, 249
Interaction of electrons with matter, 299
Interaction of X-rays with matter, 626
Interatomic covariance, 411
Interatomic vectors, 65
Interband excitation, 300
Interference function, 65
spherical, 261
Interferometry, neutron, 660
Intermolecular correlations, 525
Intermolecular force fields, 458
Internal modes, 486
Interpolation, 368
Interpolation formula, 47
Interpolation kernel, 92
Interpolation of trigonometric functions, 478
Interpretation of E maps, 232
Intraband excitation, 300
Intramolecular energy terms, 455
Intrinsic component of translation part of space-group operation, 116
Invariance of L^2, 37
Inverse Fourier transform, 8
Inverse Fourier transformation, 36, 41
Inverse rotation operator, 114
Ionic crystal, electrostatic energy of, 449
Irreducible matrix group, 176
Irreducible representations (irreps), 175
type of, 181
Ising model, 68
Isometry, 37
Isometry property, 37
Isomorphism, 176, 290–291
lack of, 255
Isomorphous addition, 283
Isomorphous crystals, 283
Isomorphous differences, 288, 291
Isomorphous heavy-atom derivatives, 580
Isomorphous replacement, 251, 282
difference Patterson functions, 251, 253
multiple, 290
single, 252, 283
techniques, integration of direct methods with, 236
Isomorphous synthesis, 283
Isotropic harmonic oscillator, three-dimensional, 17
Isotropic temperature factors, 73
Isotropy subgroups, 68, 72
Iterative low-pass filtering, 575

Jacobians, 33, 51
Joint probability distribution of structure factors, 105
Juxtaposition of chains, 578

k vector
general, 178
special, 178
uni-arm, 181
k-vector type, 180
Kernels, 58
Cauchy, 46
de la Vallée Poussin, 46
Dirichlet, 46
Fejér, 46, 464
interpolation, 92
of Fourier transformations, 35
Poisson, 46
spherical Dirichlet, 64, 91
spherical Fejér, 64
Kikuchi-line contrast, 543
Kikuchi-line form factor, 543
Kinematical approximation, 62, 300–301, 359, 394, 583, 658
Kinematical diffraction formulae, 302
Kinematical diffraction intensities, 302
Kinematical R factor, 584
Kinematical scattering, 300
Klug peaks, 267
Known structural fragment, use of, 272, 389
Kronecker symbol, 5
'*Kubic harmonics*', 12

L^p spaces, 26
LACBED (large-angle convergent-beam electron diffraction), 345
Lack of isomorphism, 255
Lagrange multiplier, 95, 106, 414
Lagrange's theorem, 68
LALS, 578
Lamellar domains with long-range order, 505–506
Landau–Peierls effect, 551–552
Langmuir troughs, 583
Languages
computer-algebraic, 122
numerically and symbolically oriented, 117
Laplace's equation, 468
Large-angle convergent-beam electron diffraction (LACBED), 345
Large values of $\mu_o t$, 641
Larmor precession, 655
Lattice, 42
base-centred, 90
body-centred, 90
centred, 73
composite, 449, 452
direct, 2, 5, 42, 121, 177, 450
face-centred, 90
nonprimitive, 71
nonstandard, 42
nonstandard period, 43
one-dimensional, 552
period, 42, 62, 68
primitive, 71
reciprocal, 2, 5, 43, 48, 62, 121, 177
residual, 48
rhombohedral, 90
standard, 42
Lattice disorder, 572
correlated, 573
Lattice distributions, 43–44, 47–48
Lattice-dynamical model, 484
Lattice energy, 458
Coulombic, 449, 453
Lattice mode, 71
Lattice-parameter mapping, 626
Lattice plane, 2, 406
Lattice sums, 44, 460
of interacting spherical Gaussians, 471
Lattice transform, 450
Lattice-translation subgroup, 123
Lattice type, 123
Laue case, 631
Laue equations, 2
Laue groups, 115
Laue point, 630
Laue scattering, 520
Layer lines, continuous diffraction on, 575
Lead, 287
Least resolvable distance, 306
Least-squares adjustment of observed positions, 413
Least-squares approximation of trigonometric functions, 478
Least-squares determination of phases, 233
Least-squares method, multivariate, 95
Least-squares plane, 410
proper, 413
Least-squares refinement, 289
Lebesgue integral, 26
Lebesgue's theory of integration, 28
LEED (low-energy electron diffraction), 356
Left action, 68, 70, 72, 79
Left cosets, 68
Left representation, 72
Legendre polynomials, 468
Leibnitz's formula, 451
Length of a function, 26
Length of a vector, 404
Lennard–Jones potential, 490
L'Hospital's rule, 451
Libration, 18
Libration tensor, 18
Librational–librational correlations, 525
Lifshitz point, 553
Lifchitz's reformulation, 98
Lindeberg–Lévy version of the central-limit theorem, 204
Line drawings, 433
Linear absorption coefficient, 628
Linear change of variable, general, 35
Linear forms, 30
Linear functionals, 28
Linear scaling of computational cost of Ewald direct sum, 474
Linear transformation, 7
Linearity, 35, 51
Linearization formulae, 75
Linearly semidependent phases, 216
Linked-atom least-squares (*LALS*) system, 578
Liouville's theorem, 37
Liquid crystals, 547
Lissajous curve, 105
Little co-group, 178, 180
Little group, 178
Local ordering, 518
Locally integrable functions, 30
distributions associated with, 30
Locally summable function of polynomial growth, 40
Location-dependent component of translation part of space-group operation, 116
Locked rotation function, 267–268
Locked translation function, 271
Logical connectivity, 435
London dispersion interactions, 475
Lone pairs, 442

Long-range order (LRO), 505, 548
  positional, 547
Longitudinal disorder, 516
Lorentz point, 630
Low-angle scattering, 508
Low-energy conformational changes, 582
Low-energy electron diffraction (LEED), 356
LRO (long-range order), 548
  positional, 547
Lyotropic phase, 548
Lysozyme, 252

MACCS, 442
Macromolecular crystallography, 282
  direct methods in, 235
Macromolecular fibre structures, 581
Macromolecular refinement techniques, 99
Macromolecular structures, direct determination of, 583
Macroscopically spherical crystal, 467
Madelung constant, 449
Magic-integer methods, 233
Magnetic crystal structures, display of, 445
Magnetic domains, 661
Magnetic scattering, 11, 656
Main reflections, 592
Manganese, 286
Many-beam case, 629
Mapping, 25
Maschke, theorem of, 176
Mathematical theory of Fourier transformation, 24
Matrices of mixed scalar products, 8
Matrix algebra, 262
Matrix–column pair, 177
Matrix groups, 176
  completely reducible, 176
  equivalent, 176
  irreducible, 176
  reducible, 176
  unitary, 176
Matrix part, 177
Matrix representation, 114
  of discrete Fourier transform, 51
Maximum determinant rule, 228
Maximum entropy, 106, 392
Maximum-entropy distributions, 36
  of atoms, 106
Maximum-entropy methods, 102, 234
Maximum-entropy theory, 106
Maximum function, 250
Maximum likelihood, 392
Maxwell's equations, 627, 643
MBD (microbeam diffraction), 356
McMurchie–Davidson recursion, 475
MDIR (multidimensional isomorphous replacement), 579
MDKINO, 439
Mean-field theory, 549
Mean-square displacement amplitude, 443
Mean values, 408
Mechanical pressure tensor, 470
Meijer's *G* function, 201
Mercury, 287
Mercury, 443–444
Mesomorphic structures, scattering from, 547
Metric
  direct, 4
  reciprocal, 4
Metric space, 25, 27
Metric tensors, 4–5
Metrizability, 25
Metrizable topology, 28
Micelle, 547
Microanalysis, 298–299
Microbeam diffraction (MBD), 356
Microdiffraction, 544–545
MIDAS, 439
Middle of reflection domain, 631
Minima
  global, 437–438
  multiple, 437
  potential-energy, 437
Minimal domain, 179
Minimization function, 289
Minimum function, 250
MIR (multiple isomorphous replacement), 290
  phases, 259
Mirror image, 285
MM2/MMP2, 442
MMS-X, 440
Modelling of diffuse scattering, 528
Modelling transformations, 431
Modified tangent formula, 233
Modulated phases, 550
Modulated smectic-A phase, 553
Modulated smectic-C phase, 553
Modulated structure, 591
Modulation function, 591
Module, 79
  over a group ring, 24
Molbuild, 442
Molecular averaging by noncrystallographic symmetry, 92
Molecular axis, 568
Molecular dynamics, 442
Molecular-dynamics refinement, 581
Molecular envelope, 32, 65, 92–93
Molecular mechanics, 437
Molecular model building, 577
Molecular modelling, 6, 418
Molecular orientational order, 547
Molecular origin, 273
Molecular replacement, 244, 258, 272, 274, 292
  real-space, 273
Molecular rotation, 558
Molecular structure, position of a known, 270
Molecular surfaces, hydrophobic and electrostatic properties of, 438
MolXtl, 443–444
Moment-generating functions, 36, 102
Moment-generating properties, 102
  of $\tilde{\mathscr{F}}$, 67
Moments of a distribution, 102
Monochromators, 660
  polarizing, 658
Monoclinic family, 118
Monoclinic space groups, 89, 118
Mosaic crystals, 626
Mosaic model, 659
Mosaicity, 356
Motif, 42–44
Motif distribution, 63
Multicritical point, 553
Multidimensional algorithms, 58
Multidimensional Cooley–Tukey factorization, 58–59, 79
Multidimensional factorization, 58
Multidimensional Good factorization, 82
Multidimensional isomorphous replacement (MDIR), 579
Multidimensional prime factor algorithm, 59
Multidimensional structure, 437
Multigrid method, 480
Multi-index, 26, 36, 38
Multi-index notation, 26
Multiple diffuse scattering, 542
Multiple isomorphous replacement (MIR), 290
  phases, 259
Multiple minima, 437
Multiple reciprocal cell, 121
Multiple scattering, 540
Multiple simultaneous superposition, 425
Multiple-wavelength method, 293
Multiplexing, generalized, 88
Multiplexing–demultiplexing, 85
Multiplication by a monomial, 25
Multiplication of distributions, 32
Multiplicative group of units, 56
Multiplicative reindexing, 61
Multiplicity, 116, 180
Multiplicity corrections, 251
Multiplier functions, 41
Multipliers, 44
  Lagrange, 95, 106, 414
Multipoles, rotation of, 475
Multireference alignment, 382
Multi-Slater determinant wavefunction, 15
Multislice, 306, 651
  calculations, 544
  computer programs, 544
  recurrence relation, 651
Multivariate Gaussian, 44
Multivariate Hermite functions, 38, 99
Multivariate least-squares method, 95
Mutually reciprocal bases, 2–3
  computational and algebraic aspects of, 4
Mutually reciprocal triads, 2
Myoglobin, 252, 282, 287

n-dimensional embedding method, 591
n-shift rule, 98
n-torus
  nonstandard, 42
  standard, 42
Natural coordinate system, 404
Negative peaks, 252
Nematic order
  biaxial, 549
  uniaxial, 549
Nematic phase, 547, 549
Nested algorithms, 62
Nested neighbourhood principle, 223
Nesting, 61
  of Winograd small FFTs, 60
Net distortions, 517
Neutral unit cell, 461
Neutron absorption, 655
Neutron beams, focusing of, 660
Neutron crystallography, 293
Neutron diffraction, 560
  dynamical theory of, 654
Neutron flight time, 660
Neutron interferometry, 660
Neutron refraction, 654
Neutron scattering
  effect of external fields, 659
  inelastic, 488
  very small angle, 660
Neutron scattering lengths, 293
Neutron spin, 655
Neutron topography, 661
Neutrons, 293
  thermal, 293
Nodes of standing waves, 633
Non-absorbing case, 629
Non-absorbing crystals, 637–639
  comparison of dynamical and geometrical theory, 640
Nonbonded interatomic distances, 578
Noncrystalline fibres, 568, 570, 576
Noncrystallographic asymmetric unit, 253
Noncrystallographic rotation elements, translational components of, 258
Noncrystallographic symmetry, 66, 258
  apparent, 267
  molecular averaging by, 92
  phase improvement using, 273
  proper, 258
  rotational, 260
Noncrystallographic symmetry element, position of, 270–271
Non-cyclic (odd) permutation of coordinates, 118, 122
Non-ideal distributions, 203, 207
Non-ideal probability density functions, 212
  of $|E|$, 205
Non-independent variables, 199
Nonlinear transformations, 264
Nonperiodic system, 503
Nonprimitive lattice, 71
Non-spin-flip scattering lengths, 657
Nonstandard coordinates, 42
Nonstandard lattice, 42
Nonstandard *n*-torus, 42
Nonstandard period lattice, 43
Norm
  Euclidean, 25
  on a vector space, 28
Normal equations, 95
Normal matrix, 98
Normal subgroup, 68–69
Normalization constant, 290, 293

Normalized Gaussians, 472
Normalized Hermite Gaussians, 473
Normalized structure factors, 216, 231, 246
Normalizer, 69
Normed space, 28
  complete, 28
Notation, multi-index, 26
Nuclear matter, density of, 2
Numerical computation of discrete Fourier transform, 52
Numerically oriented languages, 117
Nussbaumer–Quandalle algorithm, 60

O, 441
Observation plane, 361
Observational equations, 95
Obverse setting, 121
Occupancy factors, 289
Occupied natural spin orbitals, 15
Odd (non-cyclic) permutation of coordinates, 118, 122
Offset, 53
OLEX, 443–445
One-centre orbital products, 15
One-centre terms, 15
One-dimensional lattice, 552
One-particle potential (OPP) model, 22
One-phase structure seminvariants, 231
  first rank, 229
Operational calculus, 28
Operations on distributions, 31
OPP (one-particle potential) model, 22
Optic modes, 486
Optical diffractometer, 364
Optical isomers, 286
Optical rotation, 286
Optimal defocus, 361
Optimization, 436
Orbit decomposition, 69, 72, 74–75
  formula, 69, 73
Orbit exchange, 70, 77–78
Orbit of **k**, 178
Orbital products, 15
  one-centre, 15
  two-centre, 17
Orbits, 68, 72–73
Order–disorder, 507
Order parameter, 549
Orientational disorder, 524
Origin-shift vector, 123
Origin-to-plane distance, 410, 414
Origin(s)
  allowed (permissible), 215
  definition, 231
  molecular, 273
  removal from a Patterson function, 245
  selection, 247
  specification, 215
Ornstein–Zernike correlation function, 514
ORTEP, 438
ORTEP-3 for Windows, 443–444
ORTEP-III, 443–444
ORTEX, 443–444
Orthogonal matrices, 419
  eigenvalues and eigenvectors of, 425
Orthogonalization, 262
Orthographic projection, 427–428
Orthorhombic space groups, 89, 118
Oscail X, 443–444
Overlap between two Pattersons, 260

$P_s(\mathbf{u})$ function, 256
Pair probability, 504
  conditional, 504
Pairwise sum, 449
Paley–Wiener theorem, 36, 106
Parabolic equation, 648
Parallel processing, 61
Parity of the *hkl* subset, 119
Parseval–Plancherel property, 52
Parseval–Plancherel theorem, 37, 47
Parseval's identity, 64, 67
Parseval's theorem, 35, 95, 450
  with crystallographic symmetry, 74
Partial dislocations, 558
Partial net atomic charges, 449
Partial sum of Fourier series, 46
Partially bicentric arrangement, 212
Partially crystalline fibres, 572
Partially reflected wavefield, 639
Partially transmitted wave, 639
Particle–particle particle–mesh method, 475
Patterson function(s), 64–65, 75, 244, 289, 541, 577
  anomalous-dispersion, 257
  contraints on interpretation of, 258
  cylindrically averaged, 577
  deconvolution of, 249
  difference, 253–254
  generalized, 494
  interactions in, 244
  isomorphous difference, 253
  origin removal, 245
  overlap between two, 260
  second kind, 247
  sharpened, 245–246
  superposition, 250
  symmetry of, 244
  three-dimensional, 577
Patterson map, automated search, 389
Patterson peaks, 244
Patterson search, 254
Patterson synthesis, 283, 286–287
Patterson techniques, 6, 293
Patterson vector interactions, 248
PEANUT, 443–444
Pendellösung, 303, 632, 635–636, 659–660
  spherical-wave, 641
Pendellösung distance, 632–633
Penetration depth, 638
Penrose rhomb, 610
Penrose tiling, 609
Pentagonal point groups, 354
Period decimation, 48
Period lattice, 42, 62, 68
  nonstandard, 43
Period matrix, 43
Period subdivision, 48
Periodic boundary conditions, 479, 485
Periodic continuation, 544
Periodic delta functions, 211
Periodic density function, 114
Periodic distributions, 42, 44, 62
  and Fourier series, 42
  Fourier transforms of, 42
Periodic images, 460
Periodic lamellar domains, 502
Periodic weak phase objects, 364
Periodicity, 176
  crystal, 62
Periodicity requirement, 9
Periodization, 25, 44, 53
  and sampling, duality between, 44
Periodograms, 378
Permissible origins, 215
Permissible symmetry, 114
Permutation of coordinates
  cyclic (even), 118, 122
  non-cyclic (odd), 118, 122
Permutation operators, 118
Permutation tensors, 405
Perpendicular (internal, complementary) space, 592
Perspective, 418, 426–428, 434
Perturbation theory, first-order, 411
Phase angles, 282–283
Phase change, 286
Phase circles, 283
Phase determination, 283
  ab initio, 273
  direct, 36
  direct, in electron crystallography, 388
  statistical theory of, 104
Phase-determining formulae, 221
Phase evaluation, 282, 291, 293
Phase extension, 274
Phase-grating approximation, 652
Phase improvement, 274
  using noncrystallographic symmetry, 273
Phase information, 292
  from electron micrographs, 390
Phase invariant sums, 389
Phase-object approximation, 301, 542
Phase problem, 576
  Bayesian statistical approach, 106
Phase relationships
  quartet, 225
  quintet, 227
Phase restriction, 73
Phase shift, 25, 53
Phase transformations, polytypic, 514
Phases
  assignment of one or more, 231
  best, 290
  from multiple isomorphous replacement, 259
  from single isomorphous replacement, 253
  least-squares determination of, 233
  linearly semidependent, 216
  refinement of, 233
Phason flips, 611
Phasons, 527
Phenomenological absorption coefficients, 303
PHIGS (Programmers' Hierarchical Interactive Graphics System), 419
Phonon absorption, 488
Phonon dispersion relations, 489
Phonon emission, 488
Phonon scattering, 543
Phonons, 484
Physical (external, parallel) space, 592
Picture space, 426–428, 434
Pipelining, 61
Pisot numbers, 595
Pixel, 432
Plancherel's theorem, 41
Plane of polarization, 286
Plane-wave dynamical theory, 630
Planes, 406
  Gaussian, general, 413
  least-squares, 410
  least-squares, proper, 413
Plasmon scattering, 541
Plasmons
  bulk, excitation of, 300
  surface, 300
Platon, 443–444
PLUTO, 439
Pluton, 444
Point charges, finite crystal of, 465
Point density, 610
Point-group determination by convergent-beam electron diffraction, 307
Point-group operators, 115, 123
Point-group symmetry of reciprocal lattice, 114
Point groups, 177
  average multiples for, 197
  closed, 258
  decagonal, 354
  holohedral, 179
  icosahedral, 353
  pentagonal, 354
Point multipoles, 459
Point-spread function, 366
Poisson kernel, 46
Poisson summation formula, 44
Poisson's equation, 462
Polar space, 263
Polarization, plane of, 286
Polarization response, 467
Polarization vector, 11, 655
Polarizing monochromators, 658
Polycrystal electron-diffraction patterns, 358
Polycrystalline fibres, 568, 570, 576
Polycrystalline materials, diffuse scattering from, 534
Polyhedral display of crystal structures, 443
Polymer crystallography, 567
Polymer electron crystallography, 567
Polynomial growth
  functions of, 41
  locally summable function of, 40
Polynomial transforms, 61

Polynomials
  Chinese remainder theorem for, 57, 84
  cyclotomic, 57
Polyoma virus, 274
Polypeptide fold, chemical correctness of, 273
Polytypic phase transformations, 514
Population parameter, 12
Position of a known molecular structure, 270
Positional coordinates, 437
Positional order, long-range, 547
Positive peaks, 252
Positivity criterion, 293
Potassium permanganate, 286
Potential energy of a crystal, 485
Potential-energy minima, 437
PowderCell, 444
Power spectrum, 376
Poynting vector, 644
Prediction of crystal structures, 458
Pressure tensor
  mechanical, 470
  thermodynamic, 470
Pretransitional lengthening of correlation lengths, 551
Prime factor algorithm, 52, 54
  multidimensional, 59
Primitive basis, 177
Primitive coefficients, 180
Primitive lattice, 71
Primitive root mod p, 55
Principal axes, 128
Principal central projections and sections, 66
Principal projections, 76
Principal sections and projections, 67
PRJMS, 444–445
Probability
  a posteriori, 512
  a priori, 504
Probability densities, convolution of, 102
Probability density distributions, 197
  ideal, 200
Probability density functions, 203
  non-ideal, 212
  of $|E|$, non-ideal, 205
Probability density of samples for images, 366
Probability theory, 102
  analytical methods of, 102
Probability trees, 513
Processing X-ray fibre diffraction data, 574
Product function, 250
Programmers' Hierarchical Interactive Graphics System (PHIGS), 419
Projected charge-density approximation, 305
Projection approximation, 648
Projection diffraction groups, 311
Projection matching, three-dimensional, 383
Projection operator, 649
Projection(s), 25
  and sections, duality between, 41
  and sections, principal central, 66
  bounded, 67, 92
  centrosymmetric, 251–252
  orthographic, 427–428
  principal, 76
  tilt, 389
  use in three-dimensional reconstruction, 366
Projector, 70
Prolate spheroidal wavefunctions, 39
Propagation direction, 630
Propagation equation, 626
Proper least-squares plane, 413
Proper noncrystallographic symmetry, 258
Proper rotation, 123, 128
Protein crystallography, 286
Protein crystals, 287
Protein Data Bank, 438
Proteins, electron crystallography of, 389
PRXBLD, 442
Pseudo-distances, 28
Pseudorotation, 437
Pseudotranslational symmetry, 225
Punched-card machines, 76
Pure imaginary transforms, 87
Quartet phase relationships, 225
Quasicrystal structures, display of, 445
Quasicrystals
  decagonal, 354
  diffuse scattering from, 527
  icosahedral, 352
  symmetry determination of, 352
Quasilattice, 595
Quasi-long-range order (QLRO), 548
Quasimoments, 21
Quasi-normalized structure factors, 218
Quasiperiodic order, 603
Quintet phase relationships, 227

R factors
  fibre diffraction, 582
  kinematical, 584
Rader algorithm, 52
Rader/Winograd algorithms, generalized, 90
Rader/Winograd factorization, crystallographic extension of, 82
Radial dependence of atomic electron distribution, 12
Radial functions, 13
Radiation damage, 299
Radius of integration, 261
Radon measure, 30
Radon transform, 366
Random conical tilt, 377, 381
Random copolymers, 571
Random-start method, 233
Random-walk problem, 104
  exact solution, 207
Rank of tensor, 5
Rapidly decreasing functions, 37, 39
Raster-graphics devices, 432–433
Rational approximant, 595
Ray transform, 366
Real antisymmetric transforms, 88
Real crystal, 176
Real-space averaging, 273–274
  of electron density, 259, 273
Real-space molecular replacement, 273
Real-space translation functions, 271
Real spherical harmonic functions, 12
Real symmetric transforms, 88
Real-valued transforms, 85
Real waves, 640
Reciprocal axes, 405
Reciprocal Bravais lattice, 121
Reciprocal cell, multiple, 121
Reciprocal lattice, 2, 5, 43, 48, 62, 121, 177
  point-group symmetry of, 114
  weighted, 114–115
  weighted, statistical properties of, 195
Reciprocal-lattice sum, 453
Reciprocal-lattice vectors, 450
Reciprocal metric, 4
Reciprocal space, 2, 450
  symmetry in, 119
Reciprocal-space group, 175, 179, 192
Reciprocal-space procedures, 251
Reciprocal-space representation of space groups, 114
Reciprocal sum potential, 462
Reciprocal unit-cell parameters, 4
Reciprocity, 37
Reciprocity property, 36
Reciprocity relationship, 304
Reciprocity theorem, 37, 41, 43, 63, 106
  of scattering theory, 308–309
Reconstruction
  image, 361
  single-particle, 366, 375
  three-dimensional, 366
Recursion for B-splines, 477
REDUCE, 122
Reduced orbit, 74
Reducibility of the representation, 71
Reducible matrix group, 176
Reference bases, choice of, 7
Refinement
  aspherical multipole, 459
  in single-particle reconstruction, 383
  least-squares, 289
  molecular-dynamics, 581
  of phases, 233
  restrained least-squares, 581
Reflected intensity, 640
Reflecting power, 635–636
Reflection case, 631
Reflection conditions, 73
Reflection domain, middle of, 631
Reflection geometry, 631–632
Reflection high-energy electron diffraction (RHEED), 356
Reflections
  acentric, 74
  equivalent, 285
  main, 592
  satellite, 592
  substructure, 218
  superstructure, 218
Refraction, neutron, 654
Refractive index, 300
Regularization, 34
  by convolution, 42
Reindexing
  additive, 61
  multiplicative, 61
Relationship between structure factors of symmetry-related reflections, 115
Relationships between direct and reciprocal bases, 3
Relatively prime integers, 406
Relativistic effects, 301
Representation domain, 179
Representation method, 223
Representation of space groups in reciprocal space, 114
Representation of surfaces by dots, 433
Representation of surfaces by lines, 433
Representation of surfaces by shading, 433
Representation operators, 72, 78
Representation property, 68
Representations, irreducible, 175
Representative operators of a space group, 123
Repulsion energy, 449
Residual lattice, 48
Resolution
  extended, 274
  image, 306
  instrumental, 307
Restacking, 558
Restrained least-squares refinement, 581
RHEED (reflection high-energy electron diffraction), 356
Rhombohedral lattice, 90
Riemann integral, 26
Riemann–Lebesgue lemma, 36
Right action, 68, 70
Right cosets, 69
Right representation, 68
Rigid-body motion, 18
Rigid-body superposition, 422
Rigid rotation, 8
Ring systems
  condensed, 437
  flexible, 436
  fused, 442
Rings, 442
RMBD (rocking microbeam diffraction), 356
Robertson's sorting board, 76
Rocking curve, 633, 638, 660
  width at half-height, 636
  width of, 633
Rocking microbeam diffraction (RMBD), 356
Rodrigues' formula, 468–469
Root-mean-square differences between crystal structures, 445
Root-mean-square error, 291
Rotation, 426, 429
  improper, 123, 128
  molecular, 558
  optical, 286
  proper, 123, 128
  rigid, 8
  screw, 18
Rotation axes, improper, 258
Rotation-function Eulerian space groups, 265

Rotation functions, 260
  fast, 268
  locked, 267–268
Rotation matrix, 267, 419
  trace of, 263
Rotation of multipoles, 475
Rotation operator, 6
  inverse, 114
Rotation part of space-group operation, 115
Rotation vector, 421
Rotational filtering, 365
Rotational structure (form) factor, 526
Rotational symmetry, noncrystallographic, 260
Row–column method, 58

SAD–MAD (single anomalous dispersion–multiple anomalous dispersion), 238
Saddlepoint approximation, 102–103
Saddlepoint equation, 105
Saddlepoint expansion, 104
Saddlepoint method, 36, 105
SAED (selected-area electron diffraction), 307, 584
Sampling, 25, 44
  and periodization, duality between, 44
  considerations, 99
  of Gaussian density, 479
  theorems, 65
Satellite reflections, 503, 592
Satellite tobacco necrosis virus, 259
Sayre's equation, 91, 230
Sayre's squaring method, 94
Scalar pressure, 470
Scalar products, 5, 404
  mixed, matrices of, 8
Scale, 426
Scale factors, 287
Scaling of heavy-atom-derivative data sets, 255
Scaling symmetry, 603
Scanning microbeam diffraction (SMBD), 356
Scanning transmission electron microscope (STEM), 304
Scattering
  classical Thomson, 10
  coherent, 488
  Compton, 489
  critical, 551
  diffuse, 540
  forward, 647
  from mesomorphic structures, 547
  incoherent, 488
  incoherent inelastic, 488
  inelastic, 300, 540
  inelastic neutron, 488
  inelastic X-ray, 490
  kinematical, 300
  Laue, 520
  low-angle, 508
  magnetic, 11, 656
  multiple, 540
  of neutrons by thermal vibrations, 488
  of X-rays by thermal vibrations, 487
  phonon, 543
  plasmon, 541
  thermal diffuse, 300
  X-ray, 10, 62
Scattering cross sections, 550, 654
Scattering diagrams, 651
Scattering factors
  atomic, 10, 284
  complex, 255
  dynamical, 543
  spherical atomic, 10
Scattering lengths, 654
  atomic, 11
  neutron, 293
  non-spin-flip, 657
  spin-flip, 657
Scattering matrix method, 363
Scattering operator, 14
Scattering power, 285
Scattering theory, reciprocity theorem of, 308–309
SCHAKAL, 444
Scherzer defocus, 305
  conditions, 362
Scherzer phase function, 362
Schrödinger equation, 300
Schur-Auerbach, theorem of, 176
Schur's lemma, 71, 78
Scrambling, 53
Screen coordinates, 426, 428
Screen space, 430
Screw correlations, 20
Screw rotation, 18
Screw shifts, 518
Script, 442
Search directions, 94
Second Bethe approximation, 302
Second-order Born approximation, 11
SECS, 442
Section, 25
Sections and projections, 25, 66
  duality between, 41
  principal, 67
Selected-area electron diffraction (SAED), 307, 584
Selection of origin, 247
Selection rules, 101
Self-energy terms, 449
Self-Patterson, 100
  vectors, 260
Self-rotation function, 100
Self-seeding, 583
Self-vectors, 251
Semi-direct product, 69
Semi-norm on a vector space, 28
Semi-reciprocal space, 648
Separating exponent, 479
Series-termination errors, 64, 91, 99
Seven-beam approximation, 652
Shadows, 434
Shannon interpolation, 25, 50
Shannon interpolation formula, 47, 92
Shannon interpolation theorem, 65
Shannon sampling criterion, 47, 68, 93
Shannon sampling theorem, 47, 65, 275
Shape-dependent term of Ewald sum, 469
Sharpened Patterson functions, 245–246
SHELX, 443
Shift of space-group origin, 120
Shift property, 51, 65
Short cyclic convolutions, 57
Short-range order (SRO), 505, 548
  correlation functions, 518
  in multicomponent systems, 520
  parameters, 541, 545
  Warren parameters, 520
Shubnikov groups, 308
Sign conventions for electron diffraction, 301–302
Signal-to-noise ratio in electron microscopy, 376, 385
Simulated annealing, 576
Simultaneous iterative reconstruction technique (SIRT), 370
Sine strips, 76
Single anomalous dispersion–multiple anomalous dispersion (SAD–MAD), 238
Single-crystal electron-diffraction patterns, 356
Single isomorphous replacement (SIR), 252, 283
  difference electron density, 253
  phasing, 253
Single isomorphous replacement–multiple isomorphous replacement (SIR–MIR), 236
Single-particle reconstruction, 366, 375
Singular value decomposition, 370
SIR (single isomorphous replacement), 252, 283
  difference electron density, 253
  phasing, 253
SIR–MIR (single isomorphous replacement–multiple isomorphous replacement), 236
SIRT (simultaneous iterative reconstruction technique), 370
Site-symmetry group, 180
Site-symmetry restrictions, 12
Size distribution, 504
Size effect, 521
Skew-circulant matrix, 56
Sliding filter, 436
Small-angle-scattering approximation, 300
Small values of $\mu_o t$, 641
SMB (symmetrical many-beam) method, 312
SMBD (scanning microbeam diffraction), 356
Smectic films, 554
Smectic-A phase, 547, 550–551
  modulated, 553
Smectic-B phase, 547
Smectic-C phase, 551
  modulated, 553
Smectic-D phase, 561
Smectic-F phase, 556
Smectic-I phase, 556
Smooth particle mesh Ewald method, 475
Sobolev space, 41
Software, Hall symbols in, 123
Solids
  electron band theory, 629
  theory of, 9
Solution of dynamical theory, 633
Solvable space groups, 71
Solvent flattening, 92
Solvent regions, 66
Sound velocities, 490–491
Southern bean mosaic virus, 259, 274
Space-group algorithm, 119
Space-group determination by convergent-beam electron diffraction, 307
Space-group notation, explicit-origin, 127
Space-group operation, 115
  intrinsic and location-dependent components of translation part, 116
  rotation part, 115
  translation part, 115
Space-group origin, shift of, 120
Space-group-specific Fourier summations, 116
Space-group-specific structure-factor formulae, 116
Space-group-specific symmetry factors, 114
Space-group symbols
  computer-adapted, 117, 122, 127
  explicit, 123–124
  Hall, 123, 127, 130
  Hermann–Mauguin, 119
Space-group symmetry, 576
Space-group tables, 119
Space-group types, 71
  affine, 177
  crystallographic, 177
Space groups, 71, 176, 282
  cubic, 118
  Eulerian, 263
  finite, 177
  hexagonal, 119
  holosymmetric, 180
  in reciprocal space, 162
  monoclinic, 118
  orthorhombic, 118
  reciprocal-space representation of, 114
  representative operators of, 123
  rotation-function, 264
  solvable, 71
  symmorphic, 71, 177
  tetragonal, 119
  triclinic, 118
  trigonal, 119
Spaces
  data, 426–427, 429–430
  display, 426–427, 429–431, 434
  picture, 426–428, 434
  screen, 430
Special Harker planes, 249
Special **k** vector, 178
Special position, 72
  condition, 72
Special reflection, 73
Spectrometer, triple-axis, 490
Specular reflection, 554
Spherical angles, 262
Spherical atomic scattering factor, 10
Spherical atoms, 116
Spherical coordinates, 467
Spherical Dirichlet kernel, 64, 91
Spherical Fejér kernel, 64
Spherical harmonic addition theorem, 468
Spherical harmonic expansion, atom-centred, 12
Spherical harmonic functions, real, 12
Spherical harmonics, 268

Spherical interference function, 261
Spherical polar coordinates, 263
Spherical-wave *Pendellösung*, 641
Spin-flip scattering lengths, 657
Spin orbitals, occupied natural, 15
Spiro links, 442
Spot boundaries, 575
Squarability criterion, 293
Square-integrable functions, 41
Square-summable sequences, 47
Squaring method equation, 91
SRO (short-range order), 505, 548
  correlation functions, 518
  in multicomponent systems, 520
  parameters, 541, 545
  Warren parameters, 520
Stacked transformations, 431
Standard basis of $\mathbb{R}^n$, 42
Standard coordinates, 42, 63, 71–72
Standard Gaussian function, 38, 41
Standard lattice, 42
Standard n-torus, 42
Standard uncertainty of distance from an atom to a plane, 412
Standards for graphics, 419
Standing waves, 633
  anti-nodes of, 633
  nodes of, 633
Star, arms of, 178
Star of **k**, 178
Starting models, 578
Statistical properties of the weighted reciprocal lattice, 195
Statistical theory of communication, 104
Statistical theory of phase determination, 104
Statistics
  crystallographic, 204
  structure-factor, 117
Status of centrosymmetry, 123
Steepest descents, Booth's method, 96
STEM (scanning transmission electron microscope), 304
Stereochemical information, 576
Stereoviews, 428
Stirling's formula, 106
Structural connectivity, 435
Structural similarity, 425, 445
Structure amplitude, 10
Structure determination
  by X-ray fibre diffraction analysis, 576
  using electron-diffraction data, 583
Structure-factor algebra, 75, 105–106
Structure-factor formalism, generalized, 22
Structure-factor formulae, space-group-specific, 116
Structure-factor statistics, 117
Structure factors, 6, 8, 10, 62, 282, 466
  calculation of, 73
  for one-phonon scattering, 487
  Fourier–Bessel, 569
  from model atomic parameters, 93
  geometric, 116–117, 135
  in terms of form factors, 63
  inequalities among, 221
  joint probability distribution of, 105
  normalized, 216, 231, 246
  quasi-normalized, 218
  rotational, 526
  tables of, 117, 135
  trigonometric, 116–117, 135
  trigonometric, even absolute moments of, 206
  trigonometric, moment of, 205
  unitary, 218
  via model electron-density maps, 93
Structure invariants, 216
Structure seminvariants, 216
  algebraic relationships, 228
  one-phase, 229, 231
  two-phase, 229
Structure theorem, 35
  for distributions with compact support, 43, 47
STRUPLO, 444
STRUPLO for Windows, 443–444
STRUVIR, 443–444
Sturkey's solution, 648
Subcentric arrangement, 212
Subdivision and decimation of period lattices, duality between, 48
Sublattice, 48
Subspace sectioning, 436, 441
Substitutional disorder, 545, 572
Substitutional order, 543
Substructure reflections, 218
Sum function, 250
Sum of images, 249
Summable functions, 27
Summation convention, 5
Summation problem in crystallography, 46
Superposition methods, 249
Superposition of Patterson functions, 250
Superpositions
  general, 422
  multiple simultaneous, 425
  rigid-body, 422
Superstructure reflections, 218
Support, 25
  compact, 25, 36, 44
  of a distribution, 30
  of a tensor product, 34
Support condition, 34, 44
  generalized, 34
Surface effects, 553
Surface phase, 557
Surface plasmons, 300
Surface representation in cryo-EM, 386
Surfaces
  atomic, 590, 596
  dispersion, 629, 650, 657
  fractal atomic, 597
  representation of, 433
  van der Waals, 433, 439
Sybyl, 442
Symbolic programming techniques, 114
Symbolically oriented languages, 117
Symmetric rectangular density modulation, 601
Symmetrical many-beam (SMB) method, 312
Symmetry, 176, 263, 431
  conjugate, 35, 40
  conjugate and parity-related, 85
  crystal, 68
  crystallographic, 258
  cyclic, 83
  dihedral, 84
  effects on Fourier image, 114
  helical, 101, 375, 569, 576
  helical, approximate, 570
  Hermitian, 64, 74, 85
  in Fourier space, 120
  in reciprocal space, 119
  noncrystallographic, 66, 258
  noncrystallographic, molecular averaging by, 92
  noncrystallographic, proper, 258
  noncrystallographic, rotational, 260
  of Patterson function, 244
  permissible, 114
  pseudotranslational, 225
  scaling, 603
Symmetry elements
  Eulerian, 264
  three-dimensional, 308–309
  two-dimensional, 308–310
Symmetry factors, 116
  space-group-specific, 114
  tables of, 114
Symmetry-generating algorithm, 123
Symmetry group, 176
Symmetry lines, 180
Symmetry operation, 176
Symmetry planes, 180
Symmetry points, 180
Symmetry property, 39
Symmetry-related reflections, relationship between structure factors of, 115
Symmorphic space groups, 71, 177
Synchrotron radiation, 282, 293
Systematic absences, 73, 121
Systematic errors, 408
Szegö's theorem, 44, 67, 106
Table lookup schemes, 474
Tangent formula, 223, 292
  application of, 231
  modified, 233
  weighted, 224
TDS (thermal diffuse scattering), 300, 484, 540
TEM (transmission electron microscope), 303
Temperature factors, 73, 287
  anisotropic, 73
  artificial, 94, 100
  atomic, 17
  isotropic, 73
Tempered distributions, 36, 39, 41, 47, 72
  definition and examples of, 40
  Fourier transforms of, 39–40
Tensor-algebraic formulation, 2, 5
Tensor formulation of vector product, 6
Tensor product, 27, 34, 52
  of distributions, 33
  of matrices, 51, 58, 60
  structure of, 77
  support of, 34
Tensor product B-spline, 480
Tensor product property, 35, 67, 101
  of a Fourier transform, 76
Tensors, 5
  anisotropic displacement, 6
  antisymmetric, 6
  libration, 18
  metric, 4–5
  permutation, 405
  rank of, 5
  translation, 18
  translation, libration and screw-motion, 6
Test-function spaces, 29
Test functions, 39
Tetragonal family, 118
Tetragonal space groups, 90, 119
Text processing, 122
Texture electron-diffraction patterns, 357, 394
THEED (transmission high-energy electron diffraction), 356
Theory of distributions, 25, 28
Theory of solids, 9
Thermal diffuse scattering (TDS), 300, 484, 540
Thermal fluctuations, 550
Thermal neutrons, 293
Thermal streaks, 545
Thermodynamic pressure tensor, 470
Thermotropic phase, 548
Thick crystals, 363, 638
Thin crystals, 639
  comparison of geometrical and dynamical theory, 637
Thin films, 554
  crystal defects in, 542
Thomson scattering, classical, 10
Thon rings, 378
Three-axis joystick, 430
Three-beam approximation, 652
Three-beam inversion, 652
Three-dimensional Dirac delta function, 464
Three-dimensional electron-diffraction data, 391
  from a single crystal orientation, 585
  from two crystal orientations, 585
Three-dimensional fast Fourier transform (3DFFT), 475–476
Three-dimensional Hermite polynomials, 21
Three-dimensional isotropic harmonic oscillator, 17
Three-dimensional Patterson function, 577
Three-dimensional projection matching, 383
Three-dimensional reconstruction, 366
Three-dimensional structure determination by electron crystallography, 391
Three-dimensional symmetry elements, 308–309
Three-generator symbol, 123
Through-focus series method, 363
Tie point, 628
Tilt projections, 389
Tilted hexatic phase, 556
Toeplitz–Carathéodory–Herglotz theorem, 45
Toeplitz determinants, 45, 67
Toeplitz forms, 44, 67
  asymptotic distribution of eigenvalues of, 45, 68

Toeplitz matrices, 45
Tomography, electron, 377, 382
Topography, neutron, 661
Topological dual, 30, 39–40
Topological vector spaces, 28
Topology, 27, 39
  general, 27
  in function spaces, 27
  metrizable, 28
  not metrizable, 30
  on $\mathscr{D}(\Omega)$, 29
  on $\mathscr{D}_k(\Omega)$, 29
  on $\mathscr{E}(\Omega)$, 29
Topology analysis, 445
Torsion angles, 406
Total cross section, 11
Total-reflection domain, 638
  width of, 633, 638
Trace of rotation matrix, 263
Transfer function, 52
  of lens, 304
Transformation properties of direct and reciprocal base vectors and lattice-point coordinates, 116
Transformations
  affine, 120
  compound, 429
  direct-space, 120
  general, 423
  linear, 7
  modelling, 431
  nonlinear, 264
  of coordinates, 5, 7, 33
  stacked, 431
  to a Cartesian system, 3
  unitary, 37
  viewing, 426–427, 429–430, 432
  viewport, 426, 428
  windowing, 426
Transformed variance–covariance matrix, 407
Transforms
  complex antisymmetric, 87
  complex symmetric, 87
  Hankel, 101
  Hermitian-antisymmetric, 87
  of an axially periodic fibre, 101
  of delta functions, 40
  polynomial, 61
  pure imaginary, 87
  Radon, 366
  ray, 366
  real antisymmetric, 88
  real symmetric, 88
  real-valued, 85
Translate, 26
Translation, 18, 25, 426, 429
  part of space-group operation, 115
  part of space-group operation, intrinsic and location-dependent components of, 116
Translation, libration and screw-motion tensors, 6
Translation functions, 100, 269
  general, 269
  locked, 271
  real-space, 271
Translation subgroup, 177
Translation tensor, 18
Translation vector, 258
Translational components of noncrystallographic rotation elements, 258
Translational displacement, 18
Translational invariance, 650
Translations, conversion to phase shifts, 35
Transmission case, 631
Transmission electron microscope (TEM), 303
Transmission geometry, 631–633
  intensities of plane waves in, 634
Transmission high-energy electron diffraction (THEED), 356
Transposition formula, 80–81
  for intermediate results, 77
Triads, mutually reciprocal, 2
Triangular inequality, 27
Triclinic space groups, 89, 118
Trigonal space groups, 90, 119
Trigonometric functions
  B-spline approximation of, 477
  interpolation of, 478
  least-squares approximation of, 478
Trigonometric moment problem, 44
Trigonometric structure-factor expressions, vectors of, 104
Trigonometric structure factors, 116–117, 135
  even absolute moments of, 206
  moment of, 205
Triple-axis spectrometer, 490
Triple point, 552
Triplet relationships using structural information, 224
Triplets, search of, 231
Triply periodic, 627
Tunability, 294
Twiddle factors, 53, 58, 60, 62, 80
Two-beam approximation, 302, 649
Two-beam case, 629
Two-beam dynamical approximation, 394
Two-beam dynamical diffraction formulae, 303
Two-centre orbital products, 17
Two-centre terms, 15
Two-dimensional disorder, 514
Two-dimensional hexatic phase, 555
Two-dimensional symmetry elements, 308–310
Two-phase structure seminvariants, first rank, 229
Two-wavelength method, 293
Type of irreps, 181
Type of rotation (proper or improper), 123

Uni-arm **k** vector, 181
Uniaxial nematic order, 549
Uniformity, assumption of, 203, 209
Uniformizable space, 28
Unit cell, 178–179, 460
  dipole moment of, 461
  neutral, 461
Unit-cell parameters
  direct, 4
  reciprocal, 4
Unit cube, 42
Unitary matrix group, 176
Unitary structure factors, 218
Unitary transformations, 37
Unscrambling, 81
Uranium, 287

Valence density, 12
van der Waals surfaces, 433, 439
Van Hove correlation functions, 489
Variance, 407
Variance–covariance matrix, 406
  transformed, 407
Variances, 411–412
Vector interactions in a Patterson map, 248
Vector lattice, 177
Vector machines, 432
Vector map, 244
Vector overlap, 251
Vector processing, 61
Vector product, 405
  components of, 405
  tensor formulation of, 6
Vector radix Cooley–Tukey algorithm, 58
Vector radix FFT algorithms, 59
Vector relationships, 405
Vector-search procedures, 251
Vector space
  complete, 26
  norm on, 28
  semi-norm on, 28
  topological, 28
Vectors
  angle between two, 404
  components of, 5
  cross-Patterson, 260
  four-dimensional, 424
  interatomic, 65
  length of, 404
  of trigonometric structure-factor expressions, 104
  origin-shift, 123
  polarization, 11, 655
  Poynting, 644
  rotation, 421
  self-Patterson, 260
  translation, 258
VENUS, 443–445
Very small angle neutron scattering, 660
Vibrating crystals, 660
Vibrational–librational correlations, 525
Viewing transformation, 426–427, 429–430, 432
Viewport, 426, 428
Viewport transformation, 426, 428
Vitamin B_{12}, 286
Voronoi diagram, 373

Waller–Hartree formula, 541, 543
Warren short-range-order parameters, 520
Wavefield, 628, 630
Wavefunctions, prolate spheroidal, 39
Wavelengths, 299
Wavevectors, 635
Weak enantiomorphic images, 247
Weak-phase-object approximation, 304
Weak phase objects, 361, 583
  periodic, 364
Weighted difference map, 97, 99
Weighted lattice distribution, 44
Weighted reciprocal lattice, 114–115
  statistical properties of, 195
Weighted reciprocal-lattice distribution, 62
Weighted tangent formula, 224
Weighting factor, 289
Weights
  anisotropic, 413
  Gaussian, 413
Whole pattern (WP), 310
Width of rocking curve, 633
  at half-height, 636
Width of total-reflection domain, 633, 638
Wigner potential, 466
Wigner–Seitz cell, 178
Wilson plot, 287
Window, 426, 429, 434
Windowing, 426, 428
Windowing transformation, 426
Wing, 184
Winograd algorithms, 52, 56
Winograd small FFT algorithms, 57
Winograd small FFTs, nesting of, 60
Wintgen letter, 180
Wintgen position, 180
Wintgen symbol, 180
World coordinates, 426
WP (whole pattern), 310
Wyckoff letter, 180
Wyckoff position, 115, 180
Wyckoff symbols, 72

XmLmctep, 443–444
X-ray analysis, 287
X-ray fibre diffraction analysis, 567
  data processing, 574
  structure determination by, 576
X-ray scattering, 62, 293
  cross section, 550
  elastic component, 10
  inelastic component, 10
X-ray topographs, 626
X-rays, 293
  diffuse scattering of, 492
  interaction with matter, 626
X-Seed, 443–444
Xtal-3D, 444–445
XtalDraw, 443–444

z buffer, 433
ZAP (zone-axis pattern), 310
Zero-absorption case, 633
Zeroth-order Laue zone (ZOLZ) reflections, 308, 324
Zonal data sets
  view down the chain axis, 584
  view onto the chain axes, 585
Zone-axis pattern (ZAP), 310
Zones and rows, average intensity of, 196